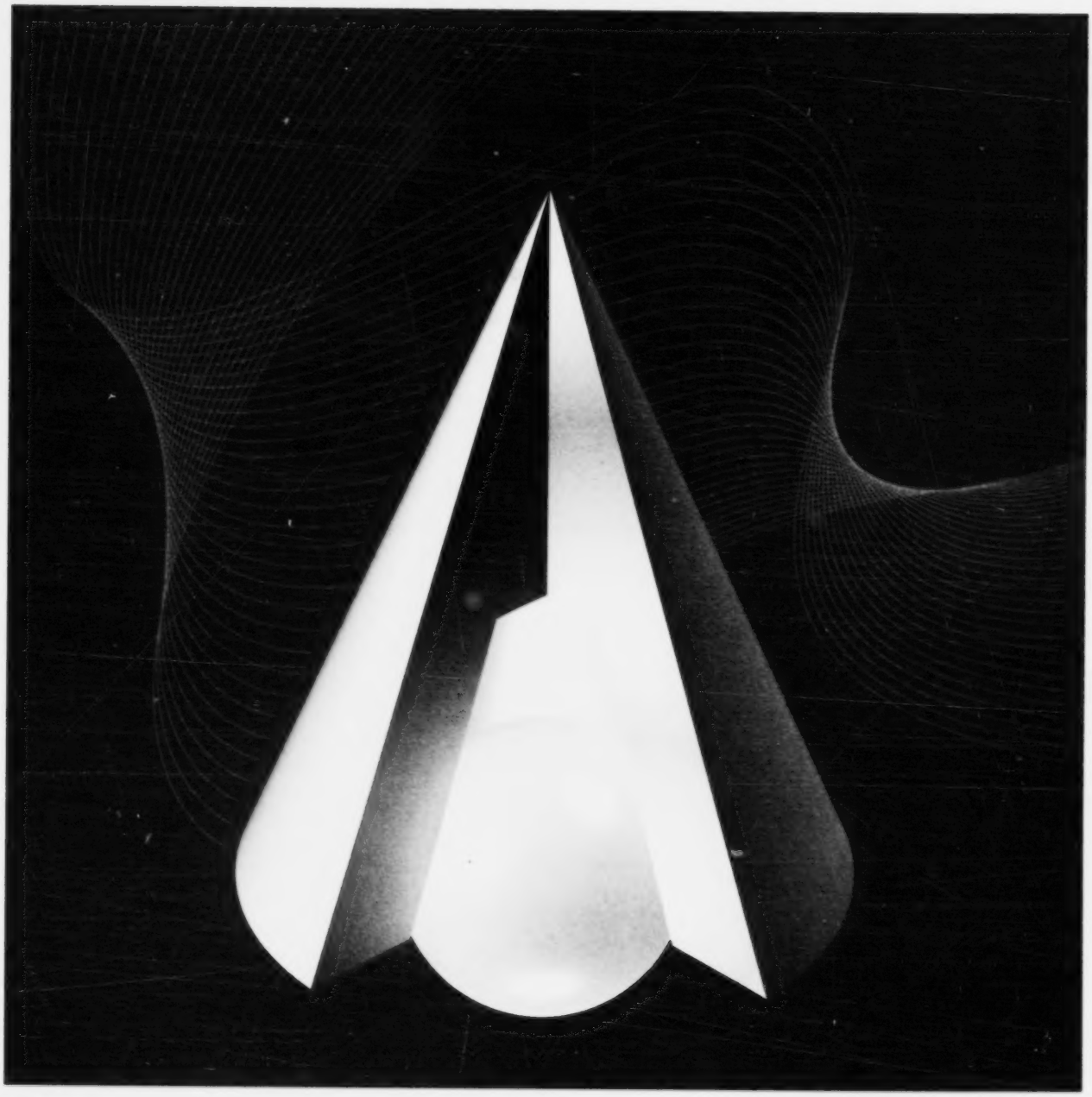
Research Paper Series

Analytical Studies

The New Migration: Global Labour Markets, Return and Onward Migration

by Abdurrahman Aydemir and Chris Robinson

No. 273



Statistics Canada Statistique Canada

Canada

Analytical Studies
Research Paper Series

The Analytical Studies Research Paper series provides for the circulation, on a pre-publication basis, of research conducted by Branch staff, visiting Fellows and academic associates. The Research Paper Series is intended to stimulate discussion on a variety of topics including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, statistical computing and simulation. Readers of the series are encouraged to contact the authors with comments, criticisms and suggestions. A list of titles appears at the end of this document.

Papers in the series are distributed to research institutes and specialty libraries. These papers can be downloaded from the Internet at www.statcan.ca

Publications Review Committee
Analytical Studies, Statistics Canada
24th Floor, R.-H.-Coats Building
Ottawa, Ontario, K1A 0T6
(613) 951-1804
(613) 951-1667

106-03698/

Return and Onward Migration among Working Age Men

by Abdurrahman Aydemir* and Chris Robinson**

11F0019 No. 273
ISSN: 1205-9153
ISBN: 0-662-42604-5

*Family and Labour Studies
Statistics Canada
Ottawa K1A 0T6

**CIBC Chair in Human Capital and
Productivity at the
University of Western Ontario

How to obtain more information:
National inquiries line:1 800 263-1136
E-Mail inquiries: infostats@statcan.ca

The paper is available on Internet: (www.statcan.ca)

March 2006

The authors wish to thank Audra Bowlus, Miles Corak, Garnett Picot and participants at a workshop on immigration issues held at Family and Labour Division in March 2005 for their helpful discussions. This project is part of the research program of the Family and Labour Studies Division, Statistics Canada.

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2006

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

Cette publication est disponible en français.

Note of appreciation:
Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Table of Contents

1. Introduction 5

2 Previous research on return migration 8

3. How permanent is the new migration? 10

4. Determinants of the length of the residence before the first departure of an immigrant after landing: Return and onward migration 11

5. Estimates of immigrant retention: Comparison with Census methods. 18

6. Conclusions and future work 21

Tables 23

Figures 32

Appendix 47

References 49

Abstract

Recent immigration appears to be characterized by frequent return and onward migration. This has important consequences for the contribution of immigrants to the economy of the host country. Lack of longitudinal data has prevented much analysis of whether recent international migration is more like internal migration and not a once-for-all move with a possible return should the move prove to have been a mistake. A newly available longitudinal data set covering all immigrants to Canada since 1980 provides the opportunity to address the issues raised by the new migration. The results show that a large fraction of male immigrants who are working age, especially among skilled workers and entrepreneurs, are highly internationally mobile.

Keywords: Immigration, return migration, visa category

JEL code: J61, J11, J68

1. Introduction

Immigration is an important issue in many developed countries. In recent policy debates, two issues are often prominent in the discussion. First is the role that immigration can play in avoiding population decline or stagnation which is implied by the low fertility rates in developed countries. The total fertility estimates for 2002 are well below replacement for many developed countries such as Australia (1.77), Canada (1.60), Germany (1.39), France (1.74), Italy (1.19), Japan (1.42), Sweden (1.54) and the United Kingdom (1.73) and approximately equal to it for the United States (2.07).[1] Immigration is a possible source of population increase to make up for the low domestic fertility rates both immediately in the form of the new immigrants themselves and in the future from the typically higher fertility rates among immigrant populations compared to native born in the developed countries. This role of immigration has received considerable attention.[2]

The second issue is the role selective immigration can play in raising living standards in the host country by increasing the supply of highly skilled workers. It is often claimed that Canada faces a brain drain of emigrants to the United States, and that skilled immigration can more than make up for this. The United States also has a large number of highly skilled immigrants arriving each year, though there is considerable debate over the average skill level of immigrants in recent years. It is generally recognized that immigrants are not randomly selected individuals from their countries of origin. They differ from non-migrants in terms of both observed and unobserved characteristics. These selection effects come from the behaviour of the migrants themselves and on the behaviour of the host country in the selectivity implied by its immigration rules.

The contribution immigrants make to the host country in either of these roles depends both on the numbers and skill levels of immigrants that come into the host country—an issue that has been studied extensively; on how long they stay—an issue that has received less attention; and on who stays—an issue that has received attention only recently. However, the issue of return or onward migration, and particularly who stays, is increasingly recognized as an important issue requiring further study.[3] It is important because it can have a major impact on the net addition made to a host country's population by immigration.[4] It also affects, via the selective nature of the process, the quality of the immigrant stock and ignoring it results in substantial biases in studies of immigrant assimilation.[5]

In addition, evidence on out-migration is important for the design of immigration policy and has implications for the payoff to the costs incurred for settlement and assimilation. Canada and the United States, for example, are major host countries and incur settlement and assimilation costs for particular classes of immigrants. To the extent that large numbers of immigrants return to their country of origin or use the initial host country as a stepping stone to another, the return to these

1. Source: *The World Factbook*, Washington DC, Central Intelligence Agency 2002; Bartleby.com, 2002.
2. See, for example, Beach, Green and Reitz (2003).
3. While the importance of assessing return migration is increasingly recognized, the phenomenon itself has a long history. Piore (1979), for example, provides estimates of large return migration flows from the United States in the early part of the 20th century.
4. Warren and Peck (1980) drew attention to the importance of the magnitude of return migration for an accurate picture of the net addition made to the United States population by immigrants.
5. See Jasso and Rosenzweig (1982) and Borjas and Bratsberg (1996).

costs will be reduced if they have used those services. If immigration policy is designed to attract permanent immigrants, it is important to understand the determinants of return or onward migration. Evidence on trends in out-migration is essential to keep policy up to date. The literature on return migration has raised awareness that migration is not necessarily a permanent move for many migrants. However, return migration itself has often been taken as permanent, if only because of the data limitations in treating it differently. In the increasingly global labour market it may be more appropriate to treat international migration more like internal migration. Individuals may move around from place to place for job-related or other reasons several times in a lifetime. Barriers to international labour movement have been reduced in recent years. In North America, the NAFTA (North American Free Trade Agreement) provisions have made some type of movement much easier. There has been considerable debate about whether it has stimulated a brain drain of Canadians to the United States, though the literature contains no evidence on whether this is permanent, or part of an increased flow back and forth. An important new data set for Mexico (the Mexican Migration Project) has stimulated a literature that examines back and forth movement from a group of Mexican villages to the United States (Massey et. al. (2002), Munshi (2003), Colussi (2004) and Angelucci (2003)).

The previous literature on return migration, briefly reviewed below, has already provided evidence of the total magnitude of return migration in several countries and a start has been made on modeling the process and testing hypotheses regarding the important determinants. There have been, however, large changes in immigration patterns, particularly in the source country patterns for migration to developed countries such as Canada and the United States. Changes over time in the characteristics of immigrants and the speed of their assimilation have been the subject of much debate, but despite the strong connection, changes in the make-up of return or onward migration have not been investigated. The lack of data has also prevented much analysis of whether international migration is increasingly more like internal migration and not a once-for-all move with possibly a return should the move prove to have been a mistake. A newly available tax-based longitudinal data set, covering immigrants to Canada since 1980, provides the opportunity to address the issues raised by the nature of the new migration.

The analysis in the paper uses two different methods to study return and onward migration. The first method uses landings records that record all immigrant arrivals to Canada, and Canadian Censuses that provide information on the number and characteristics of immigrants at a point in time after their arrival. Based on the information provided by these repeated cross-sections, out migration rates and variation in these rates by characteristics such as country of origin, the macroeconomic environment at arrival are investigated. The second method uses the landings records and the longitudinal tax filing information and infers out migration by long term absences from the tax files. The two methods provide very similar estimates of out migration rates and same qualitative results regarding variation in rates by country of origin and phase of the business cycle at arrival. This suggests that a substantial part of the absences of immigrants from the tax records is associated with not being in the country.

The annual frequency of tax filing information as opposed to Census information that is available every 5 years also allows a finer analysis of the time path of these absences, and its association with important immigrant characteristics such as visa class and language ability at arrival which is not available in Census files. The longitudinal approach using tax files provide evidence on the factors

that determine how long immigrants remain in Canada in their first spell in the country, and what happens thereafter. The length of the first spell is particularly important to assess the total or life-cycle contribution of a new immigrant arrival to the population or labour force. In addition, the longitudinal nature of the data set permits an investigation of the intermittent nature of immigrant residence in the host country. Evidence is presented on the extent of multiple moves among immigrants in Canada suggesting that neither initial nor return migration is permanent, but both are the kind of "temporary" phenomena observed in worker movement across job locations in internal migration. The data set covers a long enough time span to assess changes over time in the selectivity of the possibly temporary return or onward migration, as it relates to the global labour market.

The analysis in the paper focuses on a group of young males who are of working age. Studying this group has several advantages such as typically strong labour force attachment and high capture rates in the tax files and issues that arise when studying the entire immigrant population such as mortality are less important in this context. However, this group is also more likely to be a highly mobile segment of the population and, therefore, the results should not be generalized to the entire immigrant population.

The plan of the paper is as follows. In Section 2, the previous research on return migration is briefly reviewed. It highlights the importance of return migration and the variation in return migration by source country documented in the previous literature. Section 3 uses the longitudinal tax data to assess the intermittent nature of filing, as a proxy for residence in the host country, of immigrants arriving since 1980. It documents the temporary nature of many immigrant periods of residence in the host country. In Section 4, a variety of evidence is presented on the factors affecting the length of the first spell of residence in the host country, or conversely, the extent of return or onward migration. This evidence shows a large amount of return or onward migration and substantial variation in magnitudes over time and by various characteristics including class of immigrant and source country. For the cohorts landing around the 1990/91 recession, a substantial fraction left the country within a relatively short period of time. The migrants from source countries like Hong Kong or the United States had particularly short stays, as did those entering under the business class or skilled class category for source countries in general.

Section 4 also examines the time path of the exits, estimating hazard functions under a variety of specifications. There is a very clear pattern of particularly high hazard rates in the first year that subsequently fall rapidly to quite low levels, except for migrants from the United States. This occurs for all visa classes, indicating that most of the variation in length of stay is due to the differences in the hazard rates in the first year. The first year hazard rate for the business class, for example, is 0.311 compared to 0.167 for refugees and 0.206 for the family class. However, after this period, the hazards are all very similar; the gaps are only 0.01 to 0.02 between classes. The different shape of the hazard for migrants from the United States suggests that, not surprisingly, this particular international migration is most like internal migration. The hazard functions for the migrants from Hong Kong reflect the very strong influence of the handover to China. In particular, the hazard function for the 1980 to 1984 cohorts, which migrated largely before the handover discussions is very similar to those for other source countries outside of North America. However, for the 1990 to 1994 cohort there is a dramatic increase in the first period hazard relative to the earlier cohorts.

The results in Sections 3 and 4 are derived from tax filing behaviour rather than direct evidence on residence. Tax filing behaviour is of great interest in itself in that it allows the assessment of the life-cycle contribution of newly landed immigrants to the labour market and tax payments. However, the relation to residence is also of interest, and Section 5 presents a comparison with a census based approach that directly measures residence. The census based approach is much more limited in what can be done because of its repeated cross-sectional, rather than longitudinal nature, and the lack of information on characteristics such as visa status. However, using a synthetic cohort approach with the census provides an alternative way of estimating a subset of the results obtained in Section 4 using tax filing behaviour. A comparison of these estimates provides substantial corroboration of the results reported in Section 4. Some conclusions and an outline of future work are given in Section 6.

2. Previous research on return migration

Empirical studies of out-migration of immigrants have been hampered by the lack of longitudinal data on immigrants that would directly identify leavers. Many studies use repeated cross-sectional data, such as a national census, and focus on obtaining estimates of the amount of out-migration. Warren and Peck (1980) for example, use the U.S. censuses for 1960 and 1970, together with Immigration and Naturalization Service (INS) statistics on aliens admitted for permanent residence, to estimate total emigration in the period 1960 to 1970, and the fraction of immigrants admitted between 1960 and 1970 that had emigrated by 1970. Their estimates show that more than one million foreign-born persons left the United States during the decade. They conclude that the "implications of substantial foreign-born emigration for United States population growth are obvious. Rather than 400,000 persons being added to the United States population each year (the level of net immigration currently used by the Census Bureau in its population projections), the real addition is probably closer to 250,000 each year."[6]

A related census based approach was used for Canada by Lam (1994). This method relies entirely on census data, using a synthetic cohort approach. The estimates, based on the micro data files for the censuses of 1971 and 1981, show a substantial amount of return or onward migration. In addition, by using the individual characteristics available in both censuses, the covariates associated with return migration were investigated. A result in common with the literature for the United States is the substantial variation by country of origin.

Jasso and Rosenzweig (1982) were able to use the U. S. Alien Address Report Program which simulates a longitudinal research design. Combining this with mortality records and survey data, Jasso and Rosensweig (1982) obtain estimates of cumulative net rates of emigration for the 1971 legal immigrant cohort at about eight years after entry. An important feature of these estimates is that they were obtained by country of origin which permits some consideration of some, possibly very important, selection effects in emigration. Like the earlier literature, Jasso and Rosenzweig (1982) estimate large emigration rates: "The emigration rate for the entire cohort could have been as high as 50%. Canadian emigration was probably between 51% and 55%. Emigration rates for legal immigrants from Central America, the Caribbean (excluding Cuba), and South America were at least as high as 50% and could have been as high as 70%. On the other hand, emigration rates for Koreans

6. Warren and Peck (1980), p. 79.

and Chinese could not have exceeded 22%."[7] Borjas and Bratsberg (1996) report a similar pattern of out-migration rates by country of origin.

The major disadvantage of the methods used in this literature to estimate return migration is that it cannot examine migration at the individual level because of the reliance on the census data to identify leavers. Since there is no individual link from the administrative data to the census, individual characteristics of the leavers cannot be identified, only averages. Other disadvantages follow from the fact that the absence of this link requires a variety of adjustments to be made to the census figures to make sure that they are comparable to the administrative records cohort. These include census enumeration problems, illegal immigrants, mortality issues and census respondent recall of their immigration date many years after the fact. The impossibility of an individual level analysis from this method means that many important questions regarding the contribution of immigrants cannot be answered.

The Mexican Migration Project (MMP) provides longitudinal data to examine the patterns of movement of individuals between a group of Mexican villages and the United States. An early example of the use of this data set is Massey, Durand and Malone (2002). This data set has provided a great deal of information on the temporary nature of much of this migration, and the information on potential wages in both locations over time has presented an opportunity to model this back and forth movement. Recent papers by Munshi (2003), Colussi (2004) and Angelucci (2003) are examples of this modeling effort. The MMP data are useful for understanding a particular example of back and forth international migration for the Mexican case, but is limited to a particular source country. In addition, it cannot provide a picture of return or onward migration in total for the host country.

Constant and Massey (2002) examine return migration using the German Socio-Economic Panel, which provides a source of longitudinal data, beginning in 1984 on about 3,000 legal immigrants. This study focuses on examining selectivity in return migration and provides evidence on the nature of this selectivity. Like the census based studies in Canada and the United States, country of origin is very important. The largest immigrant population is from Turkey and these immigrants were much less likely to return than immigrants from the European Union. Studies based on national longitudinal panels can provide information on return and repeat migration that is not possible with the census based studies. They provide clear evidence that return and repeat migration are important phenomena. However, they are typically limited by relatively small sample sizes for immigrant populations, especially at a disaggregated level.

7. Jasso and Rosenzweig (1982), p. 289.

3. How permanent is the new migration?

The data sets that form the primary basis for this section and section 4 are the Landings Records (LIDS) and the Longitudinal Immigration Data Base (IMDB). The LIDS file is a rich source of immigration data, recording all landings in Canada from 1980 onwards and containing a wide variety of personal, demographic and program data including the immigrant category. The IMDB matches the LIDS with information from the tax records, thereby providing a longitudinal earnings record for immigrants that remain in Canada after landing.[8] Given that about 10 to 15 % of immigrants in the age group we are interested in never appear in the IMDB, the information in the LIDS is important to get a full count of immigrants. The longitudinal aspect of IMDB is especially valuable for a variety of important immigration related questions. In this section, the LIDS and tax data are used to examine the potentially intermittent nature of the residence of immigrants to a host country.

The IMDB provides information on the tax behaviour of immigrants who landed since 1980.[9] However, whether they work or reside in the country in any subsequent year must be inferred from the tax records. The IMDB tax records show intermittent filing for many immigrants. The IMDB records include many immigrants who have landed and filed taxes who go on to have periods of non-filing for up to four years, and yet who subsequently recommence filing. From the IMDB it is not possible to know if these individuals left the country for a period and subsequently returned. They may have been permanently residing in the country and had intermittent periods of non-filing. This intermittent nature of tax filing is of interest in itself. If the lack of tax filing does indicate absence, as an increasingly global labour market might suggest, then it will allow the calculation of the contribution of a given cohort of immigrants to the work force. If it indicates instead just the absence of paying taxes, it will allow the calculation of the contribution of the cohort to taxes.

Table 1 compares the intermittent tax filing patterns of immigrants observed in the LIDS/IMDB, with the patterns of a sample of primarily native-born Canadians. A cohort of male immigrants, aged 25 to 30 at the time of landing, are followed over time. This sample was chosen to focus on a working-age group with strong labour market attachment to minimize the probability of having no income to report. Column (1) records the fraction absent, where absence is defined as not filing over a 10 year period (i.e., perhaps leaving within a year of landing) or filing at least once in the first five years from the base year, followed by a period of 4 or more consecutive years of non-filing (i.e., perhaps leaving within 1 to 4 years of landing.) For the immigrants aged 25 to 30 who landed in 1985, 20.2% were absent, according to this definition. For the same age group that landed in 1989, this rate had risen to 33.7%. This definition does not require the absence to be permanent (see Appendix for more details). Column (3), however, shows that a large part of this absence is long term. Column (3) reports the percentage of immigrants absent according to the above definition, and who didn't reappear within 10 years from the time of landing. This ranges from 17.7% among the 1985 landings to 24.6% for the 1989 landings.

8. To be included in the IMDB, an individual has to file at least one tax return after landing.

9. Landing refers to the process where immigrants arrive in Canada with their landings documentation that starts their permanent resident status. Temporary residents that are already residing in Canada and are accepted as immigrants have to leave the country and re-enter with their landing documents for their permanent resident status to take effect. Thus, for some individuals permanent residence may start following a period of temporary residence.

Some indication of the extent to which the lack of tax filing indicates absence from the country may be gauged from a comparison of immigrant and native-born tax filing. The Longitudinal Administrative Databank (LAD) was used to provide the sample for the comparison group. The LAD, like the IMDB, is based on administrative tax records. To be included in the LAD, an individual must have filed a tax return at least once over the 1982 to 2002 period.[10] The LAD is a representative sample of all (ever) tax filers regardless of their immigrant status. The ideal comparison group is the native born; however, in the LAD it is not possible to identify immigrants who arrived before 1980.[11] For consistency across the 1985 to 1989 tax years that are examined in Table 1, the comparison samples all exclude recent immigrants, defined as those who arrived in the previous five years. These samples are all dominated by the native born. The short-term absence rates for these samples, reported in Table 1, column (2), are very similar in all years. In the 1985 tax year cohort, the rate is 13.9% and in 1989 it is 14.1%. These are much smaller than the short-term rates in the immigrant samples. In 1985 the immigrant rate is about 1 ½ times as large, and by 1989 it is more than twice as large. A similar pattern is observed for the long term rates. The long term rates for the native-born sample are reported in Table 1, column (4). These are less than half the rates in the immigrant sample. In fact for 1989, the rate in the immigrant sample is three times the rate in the native-born sample. This suggests that a substantial part of the absence of immigrants from the tax records is associated with not being in the country. The census based analysis in Section 5 supports this conclusion.

The final two columns of Table 1 document the intermittent tax filing behaviour of both the immigrant and native-born cohorts. Columns (5) and (6) report the percentage of the short-term absences that subsequently reappear within 10 years from the base year. In the immigrant samples the reappearance rates are significantly lower than in the native-born samples. This would be expected if a larger fraction of the immigrant absences were actual absences from the country, rather than just from the tax files. However, even in the immigrant samples there is substantial reappearance, suggesting that residence in Canada, especially for more recent immigrants, may be intermittent. Overall, the evidence in Table 1 suggests that neither the migration decision, nor the return migration decision of recent immigrants to Canada is permanent. Modeling immigrant behaviour, therefore, needs to take this into account.

4. Determinants of the length of the residence before the first departure of an immigrant after landing: Return and onward migration

The IMDB provides a unique opportunity to examine the life cycle profile of recent immigrants, regarding their residence or tax filing behaviour in Canada, in the context of a new global labour

10. The cross-sectional capture rate for males aged 25 to 30 in the LAD as a percentage of the corresponding population in year 1985 was about 87%. These capture rates have substantially increased in general for all age groups over the years and for the above-age cohort (age 40 to 45 in 2000), the capture rate was 98% in year 2000. High capture rates over subsequent years, therefore, put the LAD close to a full count of the whole population, like the LIDS being the full count of the immigrant population.

11. Earlier immigrant cohorts cannot be identified in this data. LAD with its linkage to Landings Records provides rich information on recent immigrants (those who arrived since 1980), e.g., education, visa category, however corresponding information such as education does not exist for either Canadian-born or non-recent immigrants.

market where mobility among immigrants is increasing. The evidence in Section 3 suggests that residence in Canada as indicated by tax filing behaviour, especially for recent immigrants, may be intermittent. In this section, rather than focusing on any arbitrary definition of permanent return migration, the analysis examines the determinants of the interval between landing and the *first* consecutive 4-year spell of non-filing. An individual may recommence tax filing after such a spell. The analysis is based on the tax filing behaviour, hence residence behaviour inferred from tax files may include some bias if for example some individuals do not file tax returns for 4 or more consecutive years although they are in the country. We show later in section 5, however, that the size of the bias is small.[12,13] Recent international migration is viewed as being sufficiently influenced by the global labour market to treat all "spells" in a given country analogously to job spells. In an internal migration setting, many individuals have intermittent spells in various jobs or occupations. All moves are potentially temporary and all jobs may be returned to. The focus of interest in this setting is an examination of the determinants of the spell lengths. In this section, the same approach is used with international migration.

The IMDB contains a rich set of characteristics on all immigrants such as visa class, education level and language ability. The IMDB makes it possible to examine how immigrant life cycle profiles are related to the various characteristics that are used to shape immigration policy. In addition, it is possible to distinguish immigrants according to source region and the immigrant class under which they are admitted, such as skilled worker or refugee. The evidence provided below shows that immigrants admitted from different regions and under different visa classes have very different life-cycle profiles of residence in Canada.

Interval regression analysis

This section reports the results of an interval regression analysis of the role of the covariates of interest in determining the length (in months) of the first spell of residence, as indicated by tax filing behavior. For the purpose of this section, the spells are referred to as spells of residence. As noted in the previous section, tax filing is not equivalent to residence, but a substantial portion of the absence from tax files for immigrants appears to be absence from the country. From the point of view of an immigrant's contribution to a host country, the spells of tax filing are of interest in themselves. However, the interpretation of the coefficients reported in this section as reflecting the determinants of residence spells, is subject to the caveat that the data refer to tax filing. The data are such that while the landing date is known precisely, i.e., no left censoring, the date of the end of the spell is only known within an interval because of the annual nature of tax filing, or not known at all because

12. A census-based approach in Section 5 that directly measures residence shows that non-filing behaviour studied in Sections 3 and 4 is mostly associated with absence from the country rather than being in the country but not filing. As discussed later, for example, for the 1981 cohort 1-year and 20-year Census retention estimates are 76.2 % and 67.6 %. The estimates from tax filing method for the same periods are 78.6 % and 65.5 %.

13. Defining absences based on shorter spells, such as 2 years, shows that this type of shorter absences are more likely to refer to short-term absences from the work force as more than half of the individuals that experience such spells reappear in the data. On the other hand, a spell definition longer than 4 years provides results very similar to 4-year definition as people who are absent from tax files for 4 years are very likely to be absent an additional one or more years. Increasing spell length definition also requires dropping later cohorts that were not in the country long enough, hence, loosing some information. Given very similar estimates between the Census method and tax filing method based on the 4 year definition we prefer this definition.

of right censoring.[14] Interval regression generalizes Tobit or censored regression models to include the interval data as well as the censored data. The Tobit specification relies on a distributional assumption of normality. Inspection of the distribution of length of residence in the data indicates substantial deviation from normality using months. Log months appears much closer to normal and thus was preferred in the analysis, though the qualitative results are not sensitive to this transformation.

Table 2 reports the interval regression analysis of the first spell of residence for each landing cohort from 1980 to 1996.[15] This analysis provides an aggregate picture of expected immigrant initial stay lengths over time. Residence is measured in log months. For the omitted cohort, 1980, the predicted length of the first spell is 5.75 log months, or just over 26 years. There are very large differences across landing cohorts, reaching a low of 5.27 log months, or just over 16 years, for the 1990 cohort. These differences imply substantial variation over time in the expected duration of the first stay and in the probability of a non-filing in the first 20 years, reported in the last 2 columns of Table 2. These are derived from the interval regression coefficients in Table 2, together with the normality assumption of the model for (log) months. This probability and expected length of first stay are plotted in Figures 1 and 2. The pattern suggests strong business cycle effects. The longest stay (Figure 2) and lowest probability of leaving (Figure 1) is the 1986 cohort. This cohort was sufficiently far from the business cycle trough in the 1990 to 1991 period to be largely unaffected by this recession. However, as the cohort entry date approaches the trough year, the length of stay falls and the probability of leaving rises. Once the cohort entry date is past the recession, the length of stay recovers to the mid 1980s level. The shortest stay (highest probability of leaving) is estimated for the 1990 cohort, which is particularly exposed to the recession.

Table 3 reports the estimated coefficients for the interval regression model including measures of the individual immigrant characteristics available in the IMDB. All the characteristics are entered as dummy variable sets to minimize functional form issues. The omitted category is a single individual with no post-secondary education, fluent in English, admitted under the family class, with age at landing 25 to 29 and arriving from North America in the 1980 landing cohort. This individual has a predicted first spell of 4.8 log months, or a little over 10 years. This is much shorter than the prediction for all individuals landing in 1980 reported in Table 2 largely because of relatively short stay of single individuals arriving from North America. Similarly, the non-filing probabilities are lower. However, the coefficients on the landing cohorts show a similar pattern to Table 2. To the extent that the covariates capture all other relevant characteristics of the cohorts, the cohort dummy variables reflect the effect of the different conditions in Canada that the cohorts face. In particular, they reflect economic conditions at the first few years after entry. The results for the different landing cohorts in Table 3 continue to show evidence of business cycle effects suggested in Table 2. Figures 3 and 4 plot the same probability and expected length of first stay as Figures 1 and 2. The results in Figures 3 and 4 control for covariates by using the estimates for a reference person with the same characteristics in each landing cohort. The patterns by landing cohort in Figures 3 and 4 are very similar to those in Figures 1 and 2.

14 The landings data records the day, the month and the year of arrival. Therefore, it is possible, for example, to distinguish immigrants by month of arrival.

15. See the Appendix for the exact definition of a spell.

The effects of individual characteristics, such as language, visa class on length of stay as predicted by the results of the interval regression analysis in Table 3 are plotted in Figures 14 and 15. Language fluency is important. In particular, bilingual immigrants, and those with fluency in French tend to stay around 25% shorter. Marriage effects are also significant with married immigrants having a stay that is 25% longer than the single and over 40% longer compared to widowed, divorced or separated. There are statistically significant differences by education, but the magnitudes are modest. Those with university degrees have a stay that is about 9% shorter than the reference group. The range of the age at landing in the sample is 25 to 45 years. Within this age range there are no major differences.

Canada's immigration system admits individuals on the basis of family ties, a refugee process, or through a points system in a variety of immigrant classes, each with their own criteria for admission: business class, skilled class, and assisted relative class. These classes have substantially different implications for the length of stay in Canada. The shortest stay is for those in the business class, self-employed, entrepreneur and the skilled worker class, followed by the assisted relative class. The longest stays occur for refugees and the "other" group which includes, importantly, the backlog clearance group. The business class group has a particularly short stay—almost 45% less than the family class. The skilled worker class also has a substantially shorter stay—around 26% less. This is consistent with the notion of a global labour market since these groups would be most likely to experience mobility induced by changing relative labour market conditions in various countries. This would not be the case for the refugee class and, in fact, the refugee class has stays that are 29% longer.

The return migration literature for the United States, shows strong differences by source country. This is clearly apparent in Table 3 for Canada, holding constant other important covariates. The omitted group is North America. All, except for the special case of Hong Kong and those from South and Central America, stay much longer than this group.[16] Those from Europe or the Caribbean and Guyana, for example, stay more than twice as long. Given the potential stimulus to mobility in North America from the NAFTA agreement, it would be interesting to examine changes in these effects over time. These major differences by source country suggest that any variation in the relative weights of source countries over time will have important implications for the permanence of the migration and the overall contribution of immigration to the labour market.

Duration analysis

The interval regression analysis provides estimates of the relation between various characteristics and the interval between landing and the first absence from the tax files of four or more consecutive years. In addition, it is of interest to know the time path of these absences. Whether absences reflect leaving the country or economic inactivity, information about the time path is valuable for both assessing immigrant integration and also for any potential policy interventions. A duration analysis is conducted to provide this information. The main advantage over the interval regression analysis is that it allows the relaxation of the strong distributional assumption required to provide this information in the interval regression analysis.

16. The special case of Hong Kong is examined below in more detail.

The tax filing data in the IMDB provides observations at an annual frequency. When an absence from the tax files is recorded, the date of the actual absence from residence or economic activity date is unknown. The date is known to have occurred in some interval given by the tax filing records. This presents a problem for calculating empirical survival and hazard functions using the standard Kaplan-Meier method where survival times are treated as observations on a continuous variable. Life table analysis can be used to produce empirical survival and hazard function estimates when the survival data have to be grouped into intervals. The procedure is as follows:

Let τ_i be the individual failure or censoring times aggregated into K time intervals, $I_k = [t_k, t_{k+1}),\ k = 1,2,...,K$, and let

d_k = number of failures in interval I_k

m_k = number of censored spell endings in interval I_k

N_k = number of persons at risk of failure at start of I_k

The product limit estimate of the survivor function is defined as:

$$S_k = \prod_{k=1}^{K} \left(\frac{n_k - d_k}{n_k} \right)$$

where $n_k = N_k - m_k / 2$ is the adjusted number at risk at the start of the interval.

While this procedure can deal with the fact that the survival data have to be grouped into intervals, an exact implementation requires the initial point to be the same. Using data on all immigrants arriving in December allows all the intervals to be one year. The sample size can be increased by pooling several months and, as an approximation, assuming a common arrival date. The empirical survivor function to the first absence for the December sample is given by Figure 5 which estimates a survival rate of about 78% by the end of the first year after arrival and just above 63% by 20 years after arrival. Most of the first absences take place within the first year. This is also reflected in the empirical hazard rate presented in Figure 6. The hazard rate declines sharply in the first year and thereafter it declines at a gradual rate to zero. These figures help determine the basic shape of the hazard function, facilitating the development of an appropriate specification for the proportional hazards regression model employed below.

Table 4 presents the estimates from a multivariate analysis of duration to the first absence using a discrete time (grouped data) proportional hazards regression framework. The list of individual characteristics is the same as that used for the interval regression analyses in the previous section. The duration model looks at the effects of same factors as in the interval regression case, and the omitted category is the same as that in the interval regression facilitating comparisons between the two models.

The model used is based on Prentice-Gloeckler (1978). This model, called a "complementary log-log" (cloglog) model, can be interpreted as the discrete time model corresponding to an underlying continuous time proportional hazards model. The underlying process is assumed to be continuous but the survival time data are recorded in bands (groups). Suppose that there are N individuals ($i = 1,...,N$) each entering a state (landing Canada) at time t=0. The instantaneous hazard rate

function (corresponding to the first absence of four consecutive years from the tax files) for person i at time $t>0$ is assumed to take the proportional hazards form:

$$\theta_i(t,X) = \theta_0(t)\exp(\beta' X_{it})$$

where $\theta_0(t)$ is the baseline hazard function which may take a parametric or non-parametric form, X_{it} is a vector of covariates summarizing observed differences between individuals at time t; and β is a vector of parameters to be estimated. For simplicity, assume that all intervals are of unit length (e.g., a month), so for each person i the recorded duration corresponds to the interval $[t_i - 1, t_i)$.

The individuals that are recorded as having left the state (i.e., disappeared) are identified by the censoring variable $z_i = 1$ while those that are still remaining in the state contribute to the right-censored spell data and are indicated by $z_i = 0$. The likelihood function for this problem can be written in terms of hazard functions as:

$$\log L = \sum_{i=1}^{n}\left\{ z_i \log\left\{ h_{t_i}(X_{it_i}) \prod_{s=1}^{t_i - 1}[1 - h_s(X_{is})] \right\} + (1 - z_i)\log\left\{ \prod_{s=1}^{t_i}[1 - h_s(X_{is})] \right\} \right\}$$

where the discrete time hazard in the j th interval is given by

$$h(t_j, X_{ij}) = 1 - \exp[-\exp(\beta' X_{ij} + \gamma_j)]$$

and where γ_j refers to the base line hazard.[17, 18]

Let $y_{it} = 1$ if person i exits the state during the interval $[t-1, t)$, $y_{it} = 0$ otherwise. Then the log likelihood can be re-written as:

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{t_i} \{ y_{ij} \log h_j(X_{ij}) + (1 - y_{ij}) \log[1 - h_j(X_{ij})] \}$$

The interpretation of the coefficients β_k is the proportionate change in hazard θ, given a one unit change in X_k. The exponentiated coefficients, $\exp(\beta_k)$, gives hazard ratios, allowing a comparison of hazard rates with the reference group. Given the life table estimates of high hazard rates in the first few years followed by a sharp decline, a piece-wise constant baseline hazard specification was adopted. This semi-parametric specification allows the hazard rates to vary by length of time since landing.

Some evidence on the overall fit of the cloglog model is presented in Figure 7. This figure compares the empirical survival function reported in Figure 5 and the predicted survival function estimated

---

17. The interval specific parameter γ_j may differ in each interval allowing for a non-parametric duration dependence. If several intervals are assumed to have same hazard rather than a different hazard for each interval a piece-wise constant baseline hazard is obtained. Baseline hazard can also be specified parametrically allowing a Weibull model or a nth order polynomial.

18. Note that the specification of the hazard rate implies: $\log(-\log[1 - h_j(X)] = \beta' X + \gamma_j$; hence, the name cloglog model.

from the cloglog model with the piece-wise constant baseline hazard specification. The survival functions are very similar, suggesting a good overall fit for the duration model.

Table 4 presents the estimates of the β coefficients in the cloglog model. These correspond to the same list of covariates that was examined in the interval regression, but measure the effect on the hazard of non-filing rather than the length of stay. The two models show a very similar pattern of results for the landing cohorts and other covariates. The 1990 cohort has the highest hazard; the lowest hazards are for 1986 in the 1980s and 1993 to 1995 in the 1990s. These exactly match the results for the interval regression in Table 3. While the coefficients of the two models are not directly comparable, the effects may be compared via the respective estimates of the probabilities reported in the last columns of Tables 3 and 4. The estimated probabilities are similar in both tables.

The independent variables all have similar effects in the two models. The marked differences for visa class, region of origin and language ability are all present in the duration model. In both models, the probability of non-filing is higher by about 5 percentage points for French or Bilingual compared to unilingual English. Those entering on business class visas have about a 10 percentage point higher probability of an absence in both models. The region of origin differences are slightly exaggerated in Table 4. Europe, for example, has about a 20 percentage point higher probability in Table 4 compared to about 15 points in Table 3. Overall, the patterns are very similar in both models, and the magnitudes are similar, with a tendency for a slightly lower probability in the duration model.

The time path of exits

The hazard rates from the duration model provide a picture of the time path of non-filing. The results reported above indicate that for many types of immigrant, the length of stay can be quite short, especially for those entering under the business or skilled worker class. Examination of the shape of the hazard function shows that this is primarily due to particularly high hazard rates in the first year after landing. Figures 8a and 8b plot the estimated discrete hazard and survival functions by visa class using the cloglog variant of the proportional hazards model with a piecewise linear hazard. In Figure 9, the same piecewise linear hazard functions are plotted without imposing proportionality—i.e., estimating the model separately by visa class. It is clear from both figures that, for all visa classes, the hazard rates are high in the first year, but subsequently fall rapidly to quite low levels. This indicates that most of the variation in stay length is due to the differences in the hazard rates in the first year.[19] The first period hazard rate for the business class, for example, is 0.311 compared to 0.167 for refugees and 0.206 for the family class. After this period, the hazards are all very similar; the gaps are typically only about 0.01 to 0.02 between classes. The restricted and unrestricted estimates are quite similar, though the unrestricted model shows a particularly high initial hazard for the business class and a clearer tendency for the skilled worker class to have continuing non-negligible exit for several years.

Some differences in the shape of the hazard function by source country are apparent in Figures 10a, 10b and 11. Figures 10a and 10b plot the estimated discrete hazard and survival functions by source country using the cloglog variant of the proportional hazards model with a piecewise linear hazard.

19. The estimated hazard rates are discrete, due to the nature of the data. It is not possible to estimate the path of the underlying continuous hazard rates within the 12-month intervals. Thus, unfortunately, it is not possible to be more precise about the hazard shape in the first 24 months when most of the change in the hazard takes place.

In Figure 11, the same piecewise linear hazard functions are plotted without imposing proportionality by estimating the model separately by source country. The patterns are similar in the two figures. However, it is clear from the unrestricted estimates in Figure 11 that the shape of the hazard is different for North America and that Hong Kong is an outlier in terms of the magnitude of the difference in the first year hazard values. The immigrants from the United States show a clear tendency for the hazard rates to remain quite high for many years after entry. The other source countries show a drop to very low levels after the first year. This is consistent with immigration from the United States being particularly similar to internal migration. That is, the pattern is consistent with a relatively low cost of migration and a relatively high probability of continuing movement in response to movement in wage differences or career demands.

Hong Kong is a special case because of the effects on migration to Canada due to the impending handover to China that finally took place in 1997. The right of residence in Canada became particularly attractive and many business people acquired this right through immigrating under the business class provisions. Figure 12 shows landings in Canada from Hong Kong from 1980 to 1996. It is clear that there were large increases after 1984, when the Joint Declaration was signed between Britain and China. Figure 13 compares the estimated hazard rates for the 1980–1984 and the 1990–1994 landings from Hong Kong. After the first year, the hazard rates are the same, but the first period hazard rates are dramatically different. For the 1980–1984 cohorts, the hazard function is very similar to the function for the rest of Asia (Figure 11). However, for the 1990–1994 cohorts, while the hazards look the same after the first period, the first period hazard rate for these cohorts is much higher than the rest of Asia and the earlier Hong Kong landings. This suggests that a significant fraction of these landings may have been stimulated by the attractiveness of establishing citizenship, rather than the prospect of a long term stay in Canada.

5. Estimates of immigrant retention: Comparison with Census methods

The IMDB provides information on tax filing behaviour rather than residence behaviour directly. The evidence in Section 3 suggests that the two are closely related, at least for males. In the previous literature on return migration, a major focus of interest is estimating the retention rate of immigrants to answer the question: Given that return migration takes place, what fraction stay in the host country? Census data are used in this calculation. In this section, estimates of the retention of immigrants in Canada are presented, based on a conceptual framework analogous to that of Borjas and Bratsberg (1996) for the United States. They work with a generic out-migration rate defined as:

$$q(t, t') = [I(t) - R(t')] / R(t')$$

where I(t) is the number of persons who immigrate in year t and R(t') is the number of those immigrants who remain as of t'. The source of I(t) in Borjas and Bratsberg (1996) is INS microdata which recorded every legal immigrant admitted into the U.S. between July 1, 1971 and September 30, 1986. The source of R(t') is the 1980 Census so that t' is April 1, 1980.[20] The analogous sources

20. Some adjustment was necessary to make the census comparable with the INS. In particular, an estimate of illegal immigrants in the census was necessary since the INS covers only legal immigrants. In addition, the immigrant cohorts were "aged" to April 1, 1980 using age/sex specific mortality rates to estimate survival. For a detailed description of the

for Canada are LIDS for I(t) and the relevant Canadian census for R(t'). The retention of immigrants in Canada can then be measured by fraction of immigrants arriving at time t who are still retained at time t':

$$r(t',t) = R(t')/I(t)$$

The time path of the retention percentages for males, r(t',t), for the one-year landing cohorts that match up with the Canadian census periods are given in Table 5.[21,22] The time pattern is very similar to that documented in the interval regression and duration analysis using the IMDB. The census years 1981 and 1991 were both recession years; the 1986 and 1996 years were not. Comparing each pair of years the 5-year survival rates are substantially lower for the more recent cohorts. For all males, the 1981 cohort percentage retained after 5 years is 80.9 compared to 72.6 for the 1991 cohort—a decline of 10%. The 10-year survival rates fall even more: the retained percentage for the 1981 cohort after 10 years is 77.5 compared to 64.0 for the 1991 cohort—a decline of 17%. Similarly, across the 1986 and 1996 cohorts, there is a fall in the percentage retained after 5-years from 90.2 to 76.3—a decline of 15%.

The population of male immigrants includes both workers and non-workers whose emigration rates are likely to be influenced by various factors in different ways. In the lower half of Table 5, the survival percentages are presented for males with age at landing between 25 and 35 to capture a young working age population. The patterns observed in the total population of males show yet higher mobility in this population. The 5-year survival percentages for the 1981 and 1991 cohorts are 76.2 and 63.7, respectively—a decline of 16%; after 10 years the decline is 26%. For the 1986 and 1996 cohorts the decline in the 5-year survival percentage is 25%. These are large declines in landing cohorts separated by only a decade. There is also the same business cycle pattern, as observed in the previous sections, with low retention rates for cohorts arriving in the trough census years of 1981 and 1991 on either side of the high retention rate in the peak census year of 1986.

In Section 3, source region was shown to have a very large effect on the survival rates, based on the tax filing records. Table 6 examines the census based retention rates by source region—a characteristic that is available in the census as well as the landing records. The pattern of the results is very similar to those based on the tax records. Among the groups in Table 6, the lowest rates are for North America and the highest for Africa. The sample sizes for the source country breakdown

---

adjustments, see Borjas and Bratsberg (1996), pp. 168–170.

21. This abstracts from problems of mortality and illegal immigrants dealt with in Borjas and Bratsberg (1996), focusing on trends rather than absolute rates. Implicitly it is assumed that mortality and illegal immigration rates are stable over the period.

22. A similar approach is used by Michalowski (1991) for Canada and these estimates inform demographic projections. That study, however, looks at all age groups as opposed to our focus on 25-45 year olds at landing. Also, the author mentions the problem of over counting and under counting in the Census but corrects the estimates only for under counting that tends to yield higher retention rates. The results in this paper show that the out migration rates vary with the business cycle conditions at arrival as well as country of origin and visa class. Therefore the results imply that the shifts in the country of origin and visa distribution of immigrants over time and different macro conditions are important factors for estimating out-migration rates for demographic projections. For purposes of demographic projections estimates for the entire immigrant population is needed and the tax method has certain limitations since tax capture rates for certain groups, such as children, are low. The method could be refined by defining the family as the unit of analysis to overcome some of the problems. The data available to this study doesn't have the necessary information to construct families.

are considerably smaller than for the previous table and many of the point estimates for Africa, for example, are greater than 100%. In addition to sample size, however, there are other problems with combining census and landing records data discussed above. The actual magnitudes should thus be interpreted with some caution.

Overall, the patterns observed using the census based estimates, where they can be compared, support the patterns found using the tax filing data in the IMBD. Thus, both the landing cohort and source country effects have the same effects. The IMDB permits a much richer analysis since it includes information on other important immigration related characteristics, especially visa class. To the extent that the patterns in the census and tax filing record approaches are very similar where they overlap, this lends credibility to the use of the tax record approach in analyzing the other characteristics available in the IMDB, as well as to the longitudinal analysis at the individual level which cannot be carried out with the census approach.

Table 7 reports a comparison of the tax filing based residence definition and the census based method. When the retention rates are calculated if four years of consecutive non-filing were considered equivalent to emigration, Table 7 shows that rates are close to the estimates in Table 5. The 20-year retention rate for males with landing age 25 to 35 in Table 5 is 67.6%, while the same rate in Table 7 is 65.5%. The 15-year retention rates are identical for the 1986 cohort, while the rate for 1981 cohort is about four percentage points lower in Table 7. For 1991 and 1996 the estimates from both methods show a decline from 1986, but less so in the tax filing method. The census data are affected by a change in the question in 1991. In 1991, the questions ask specifically about the year landed immigrant status was obtained whereas prior to that the question was more vague, asking in what year the person immigrated to Canada which may not always correspond to the year of landing. This may be part of the difference. However, tax filing behaviour could also change over time. During the late 1980s and early 1990s in order to get a number of government transfers, such as child tax benefits, individuals were required to file tax returns even if they had no taxes payable. There was no such requirement before this period and this might have affected tax filing behaviour.[23] The census estimates of retention rates abstract from issues of disappearance and re-appearance. They measure what fraction of a cohort is in the country at a particular point, whether they stayed there all the time, or left and re-entered. The tax-filing estimates presented in Table 7 neglect re-entry which was shown in Section 3 to be quite significant for more recent cohorts. Taking this into account may significantly increase the tax filing based retention estimates in Table 7.

Overall, the results in Table 7 suggest that tax filing behaviour is closely related to residence behaviour as reflected in estimates of retention rates. Of additional interest is how close the relationship is in terms of the shape of the hazard rate. The hazard rate estimates presented in Section 4 (Figure 5) show a very sharp drop in the first year. This results in the survival rate falling to about 78% by the end of the first year. After this, the drop is much more gradual. Using the 1991 and 1996 censuses, the same basic pattern appears with a very similar magnitude for the initial drop. For those landing in 1990, the 1991 Census records a survival rate of 79.4%; for those landing in 1995, the

23. Figure A1, in the Appendix, shows the tax filing rates by sex over the 1980–2000 period for those who were 25 to 35 years old during the tax year. The filing rates of females increased about four percentage points during the 1990s, whereas for males the effect was much lower. Thus females filing behaviour appears to be more sensitive to changes in filing incentives of the type that occurred. The restriction of the sample to males reduces the sensitivity of the estimates to time variation in tax filing behaviour.

1996 Census shows a survival rate of 82.5%.[24] Subsequent survival rate drops are much smaller. These magnitudes imply very similar hazards to those obtained from the tax filing based estimates. Thus, even for the first year after landing, where many of the disappearances take place, the tax filing behaviour appears to closely mirror the residence behaviour, as measured in the census.

6. Conclusions and future work

There is increasing evidence that international migration, like internal migration, is not a permanent move and that many immigrants either return to the source country, perhaps many times, or move on to another country. The IMDB presents an opportunity to study this phenomenon on a large group of immigrants from a wide variety of source countries over more than a 20 year period. The analysis conducted in this paper provides estimates of the extent of return and onward migrations, as well as more repeated back and forth movement. In addition, it relates the expected life-cycle residence behaviour of the cohorts of immigrants landing since 1980 to a variety of individual characteristics. Since the IMDB contains information on all the characteristics used to implement the points system used in Canada to determine eligibility for admittance, this evidence is particularly relevant for any discussion of amendments to immigration policy based on changes to the points system.

It is clear that a substantial part of migration to Canada is temporary The estimated out-migration rate 20 years after arrival is around 35 % among young working age male immigrants. About 6 out of 10 of those who leave do so within the first year of arrival which suggests that many immigrants make their decisions within a relatively short period of time after arrival. Controlling for other characteristics, the out-migration rates are higher among immigrants from source countries such as the United States and Hong Kong, and for those admitted under the skilled worker or business class visa. Among immigrants that arrived in either business class or skilled worker class about four in 10 left within 10 years after arrival. For the assisted relative class and the refugees the corresponding rates were around three in 10 and two in 10. Finally, the out-migration rates are higher for those who arrive during recessionary periods. Immigrants who arrived in 1990, for example, were about 50 % more likely to leave than those who arrived in 1986, controlling for other characteristics.

In view of the potentially temporary nature of all migration, calculation of the contribution that can be expected from a new immigrant to the population, the labour force, or the human capital stock of a country has to take into account the probability of the immigrant being in the country at each point over the immigrant's remaining lifespan. The duration analysis in Section 4 presents a start in providing the information needed for this calculation. The analysis shows that an immigrant's future profile is strongly influenced by a variety of measurable factors, such as source country and visa class. Of particular significance for the human capital stock of the country is the fact that immigrants coming under the skilled worker class or business class often leave quite soon. In addition, there appear to be strong business cycle effects. The cohorts most affected by the recession of 1990/91 had particularly early departures.

---

24 An alternative method to estimate emigration rates using the tax filing behaviour would be to use a sample of tax filers only, i.e. those that ever appear in the IMDB. About 10 % to 15 % of immigrants never appear in the tax files. If we restrict our sample to those that ever appear in the IMDB then the estimated emigration rates would be half of those implied by the Census method. This bias in estimates highlights the importance of starting with a full count of immigrants rather than using tax filers only for the analysis.

The tax filing information in the IMDB was used in this study to estimate the life-cycle profile of immigrant residence in Canada. However, the IMDB also contains accurate income data from the tax filing information. In future work, this income information will be used to estimate life-cycle paths of income and tax contributions over time for immigrants distinguishable according to the individual characteristics available in the IMDB. In addition, it will be possible to estimate a more behavioural model, allowing a fuller picture of likely changes in immigration under a variety of policy simulations. The behavioural modeling will be more complicated than that used with the Mexican Migration Project data. In the MMP, there is an approximately closed system with movement between particular Mexican villages and the United States. The opportunity set for the migrant in Mexico is characterized by conditions in particular villages. For migration from a wide range of international sources, where the departures may not necessarily be to a well-defined, known origin, the issue is more complicated. However, for some groups, the literature reviewed in this paper provides evidence that most of the departures are a return to the source country, so that a definition of the opportunity set based on general conditions in the source country may yield a good approximation.

Table 1
Intermittent tax filing behaviour of males, aged 25 to 30

Base year	Short-term absence from tax files (%)		Long-term absence from tax files (%)		Reappearance after short-term absence (%)	
	Immigrant	Native-born	Immigrant	Native-born	Immigrant	Native-born
1985	20.2	13.9	17.7	7.3	12.5	47.9
1986	20.2	13.7	15.8	7.5	21.9	45.5
1987	23.2	13.7	17.9	7.8	22.9	43.3
1988	27.0	13.8	20.6	8.1	23.7	41.6
1989	33.7	14.1	24.6	8.4	26.9	40.6

Notes: The immigrant samples are the full count of males aged 25 to 30, derived from the LIDS, that landed in the base year; the native-born samples are the sample of males aged 25 to 30 who are either native-born, or who landed 5 or more years ago, derived for each base year from the LAD sample.
Short-term absence refers to becoming non-filer as defined on page 10.
Long-term absence refers to becoming a non-filer and not reappearing in the tax files within the next 10 years from the time of landing. Columns (5) and (6) report the difference between short term and long term absences as percentage of short-term absences.

Source: Landing Records (LIDS), Immigration Database (IMDB) and Longitudinal Administrative Databank (LAD).

Table 2
Basic Interval regression model of the first spell (log months) for landing cohorts 1980 to 1996: Males, aged 25 to 45 at arrival

	Coefficient	Standard error	Expected length of month spell	Probability of absence
Constant (1980)	5.752	0.059	314	0.458
1981 cohort	0.180	0.084	376	0.430
1982 cohort	0.232	0.084	397	0.422
1983 cohort	0.330	0.097	437	0.407
1984 cohort	0.329	0.097	437	0.407
1985 cohort	0.398	0.097	468	0.397
1986 cohort	0.639	0.090	596	0.361
1987 cohort	0.338	0.079	441	0.406
1988 cohort	0.000	0.078	314	0.457
1989 cohort	-0.267	0.074	240	0.499
1990 cohort	-0.484	0.071	194	0.532
1991 cohort	-0.295	0.070	234	0.503
1992 cohort	0.100	0.071	347	0.442
1993 cohort	0.432	0.074	484	0.392
1994 cohort	0.165	0.076	371	0.432
1995 cohort	0.325	0.078	435	0.408
1996 cohort	0.102	0.075	348	0.442
Number of observations		62,745		
Log likelihood		-78424.75		

Note: The reference group is landing cohort 1980.

Source: Calculations by authors based on Landing Records (LIDS) and the Longitudinal Immigration Database (IMDB) data.

Table 3
Interval regression model of the first spell (log months) for landing cohorts 1980 to 1996: Males, aged 25 to 45 at arrival

	Coefficient	Standard error	Expected length of month spell	Probability of absence
Constant (1980)	4.815	0.087	123	0.613
1981 cohort	0.300	0.081	166	0.566
1982 cohort	0.324	0.082	170	0.562
1983 cohort	0.477	0.095	198	0.538
1984 cohort	0.409	0.094	185	0.549
1985 cohort	0.541	0.095	211	0.527
1986 cohort	0.696	0.088	247	0.502
1987 cohort	0.523	0.077	208	0.530
1988 cohort	0.282	0.076	163	0.569
1989 cohort	-0.139	0.072	107	0.634
1990 cohort	-0.273	0.070	93	0.654
1991 cohort	-0.141	0.070	107	0.634
1992 cohort	0.326	0.071	170	0.562
1993 cohort	0.698	0.074	247	0.502
1994 cohort	0.524	0.075	208	0.530
1995 cohort	0.607	0.076	226	0.517
1996 cohort	0.392	0.074	182	0.551
Non-university post-secondary	0.076	0.032	133	0.601
BA degree or above	-0.089	0.031	112	0.626
French	-0.269	0.057	94	0.654
English and French	-0.344	0.051	87	0.665
Neither English nor French	0.235	0.032	155	0.576
Married	0.287	0.029	164	0.568
Widowed, separated, divorced	-0.177	0.081	103	0.640
Business class	-0.582	0.055	68	0.699
Skilled class	-0.305	0.037	90	0.659
Assisted relative class	-0.148	0.046	106	0.635

Table 3
Interval regression model of the first spell (log months) for landing cohorts 1980 to 1996: Males, aged 25 to 45 at arrival (concluded)

	Coefficient	Standard error	Expected length of month spell	Probability of absence
Refugee class	0.256	0.041	159	0.573
Other admission category	0.166	0.053	145	0.587
Age at arrival 30 to 34 years	-0.049	0.032	117	0.620
Age at arrival 35 to 39 years	-0.074	0.037	114	0.624
Age at arrival 40 to 45 years	-0.008	0.041	122	0.614
Europe	1.009	0.066	338	0.452
Asia, excluding Hong Kong	0.740	0.065	258	0.495
Hong Kong	-0.549	0.070	71	0.694
Middle East	0.095	0.056	135	0.598
Africa	0.870	0.076	294	0.475
Caribbean and Guyana	1.241	0.080	426	0.416
South & Central America	-0.136	0.079	107	0.634
Oceania and Australia	0.521	0.123	207	0.531
Number of observations	62,722			
Log likelihood	-76860.59			

Note: The reference group is landing cohort 1980, no post-secondary education, fluent in English, single, family class, landing from North America at age 25 to 29.

Source: Calculations by authors based on the Landing Records (LIDS) and the Longitudinal Immigration Database (IMDB) data.

Table 4
Discrete time duration model for landing cohorts 1980 to 1996: Males, aged 25 to 45 at arrival

	Coefficient	Standard error	Probability of absence
1981 cohort	-0.106	0.092	0.543
1982 cohort	-0.176	0.106	0.518
1983 cohort	-0.149	0.114	0.528
1984 cohort	-0.005	0.111	0.579
1985 cohort	-0.252	0.113	0.491
1986 cohort	-0.455	0.102	0.424
1987 cohort	-0.147	0.093	0.528
1988 cohort	-0.059	0.088	0.560
1989 cohort	0.162	0.084	0.641
1990 cohort	0.175	0.082	0.646
1991 cohort	0.084	0.080	0.612
1992 cohort	-0.328	0.086	0.466
1993 cohort	-0.452	0.094	0.425
1994 cohort	-0.378	0.094	0.449
1995 cohort	-0.495	0.096	0.412
1996 cohort	-0.201	0.088	0.509
Non-university post-secondary	-0.046	0.039	0.564
BA degree or above	0.073	0.037	0.608
French	0.110	0.068	0.622
English & French	0.175	0.059	0.646
Neither English nor French	-0.154	0.039	0.526
Married	-0.193	0.033	0.512
Widowed, separated, divorced	-0.017	0.093	0.575
Business class	0.275	0.065	0.682
Skilled class	0.194	0.043	0.653
Assisted relative class	0.044	0.055	0.598

Table 4
Discrete time duration model for landing cohorts 1980 to 1996: Males, aged 25 to 45 at arrival (concluded)

	Coefficient	Standard error	Probability of absence
Refugee class	-0.144	0.050	0.529
Other admission category	-0.009	0.061	0.578
Age at arrival 30 to 34 years	0.086	0.037	0.613
Age at arrival 35 to 39 years	0.122	0.043	0.626
Age at arrival 40 to 45 years	0.099	0.049	0.618
Europe	-0.619	0.071	0.374
Asia, excluding Hong Kong	-0.471	0.068	0.419
Hong Kong	0.302	0.072	0.692
Middle East	-0.065	0.068	0.558
Africa	-0.520	0.083	0.404
Caribbean and Guyana	-0.828	0.092	0.316
South & Central America	0.112	0.083	0.623
Oceania and Australia	-0.401	0.144	0.442
Constant (1980)	-3.880	0.100	0.581
Number of observations		111,822	
Log likelihood		-14907.03	

Notes: Pooled October-December sample. The model includes a piecewise linear baseline hazard for 1–12 months, 13–24 months, 25–36 months, 37–60 months, and after 60 months.

Source: Calculations by authors based on the Landing Records (LIDS) and the Longitudinal Immigration Database (IMDB) data.

Table 5
Census based retention rates at 5, 10, 15 and 20 years after landing: Males

		Retention rates at various years after landing			
Year	Landings	5 years	10 years	15 years	20 years
All males					
1981	63,470	80.9	77.5	71.6	68.8
1986	49,380	90.22	86.6	82.7	..
1991	116,720	72.6	64.0	..	..
1996	111,290	76.3	..	..	..
Males aged 25 to 35 at landing					
1981	18,040	76.2	76.9	70.8	67.6
1986	15,580	88.6	86.3	78.5	..
1991	40,860	63.7	57.2	..	..
1996	32,920	65.8	..	..	..

Notes: The number of landings are from the Landings Records and are for the calendar year. The retention rates are based on the census counts in the relevant census years of individuals recording their year of migration.
.. not available for a specific reference period.

Source: Landing Records (LIDS) and Census data.

Table 6
Census based retention rates at 5, 10, 15 and 20 years after landing, by source region: Males aged 25 to 35 at landing

		Retention rates at various years after landing			
Year	Landings	5 years	10 years	15 years	20 years
North America					
1981	1,440	50.7	44.1	45.1	40.3
1986	750	72.0	72.0	56.7	..
1991	580	74.1	72.4	..	..
1996	690	54.3	..	..	..
Europe					
1981	7,680	74.0	68.8	61.8	59.6
1986	4,140	83.0	85.9	73.4	..
1991	8,240	70.1	59.4	..	..
1996	7,030	72.3	..	..	..
Asia					
1981	5,850	91.2	86.8	86.4	75.4
1986	6,650	87.5	91.3	83.5	..
1991	21,050	63.0	53.6	..	..
1996	19,860	58.3	..	..	..
Africa					
1981	820	99.4	116.5	122.0	113.4
1986	1,170	124.0	110.3	83.8	..
1991	4,180	69.5	54.3	..	..
1996	2,490	80.9	..	..	..

Source: Landing Records (LIDS) and Census data.

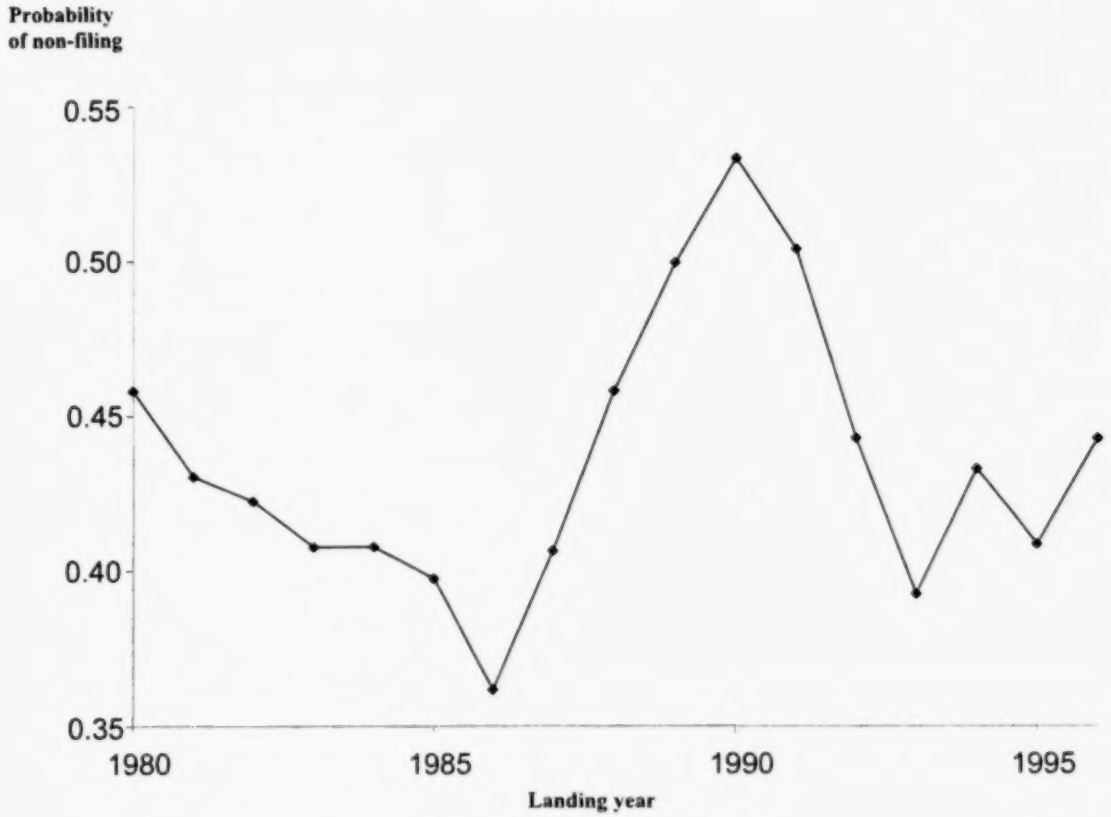
Table 7
Tax filing retention rates at 5, 10, 15 and 20 years after landing: Males aged 25 to 35 at landing

		Retention rates at various years after landing			
Year	Landings	5 years	10 years	15 years	20 years
1981	18,040	78.6	70.0	67.1	65.5
1986	15,580	84.5	81.5	78.5	..
1991	40,860	77.7	73.4	..	..
1996	32,920	77.0	..	..	..

Notes: The number of landings is from the Landings Records. The retention rates are based on the assumption that non-tax-filing in 4 consecutive years in the IMDB constitutes emigration.
.. not available for a specific reference period.

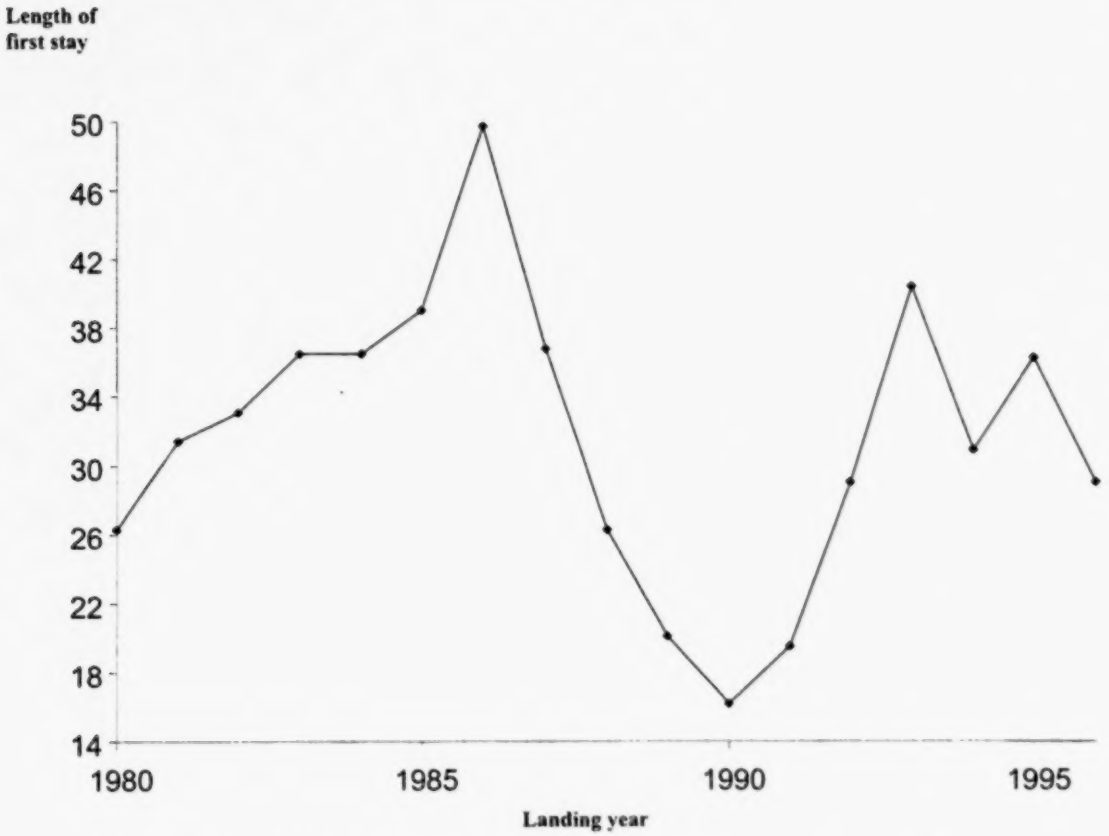
Source: Calculations by authors based on Landing Records (LIDS) and the Longitudinal Immigration Data Base (IMDB) data.

Figure 1
The probability of an absence in the first 20 years by landing year



Source: Authors' calculation based on model estimates from Table 2.

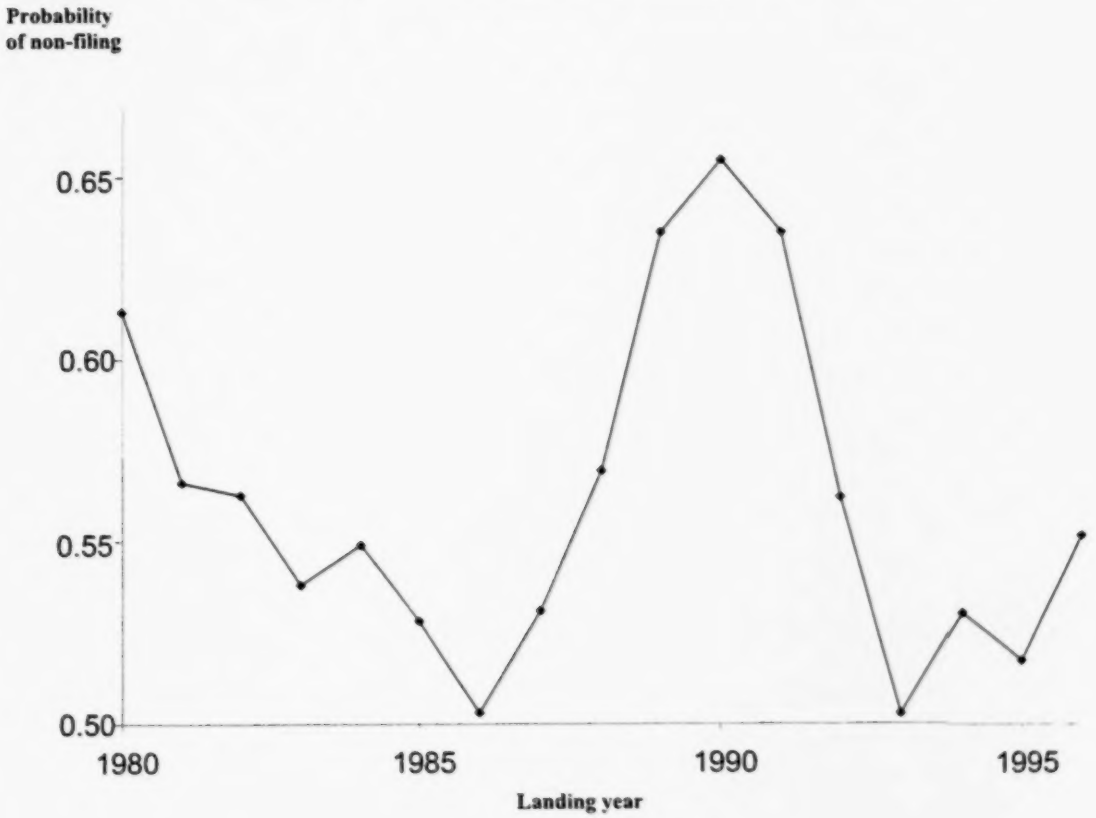
Figure 2
The expected length (in years) of the first spell by landing year



Note: The expected lengths of stay estimated in Table 2 are converted into years in this figure for expositional purposes.

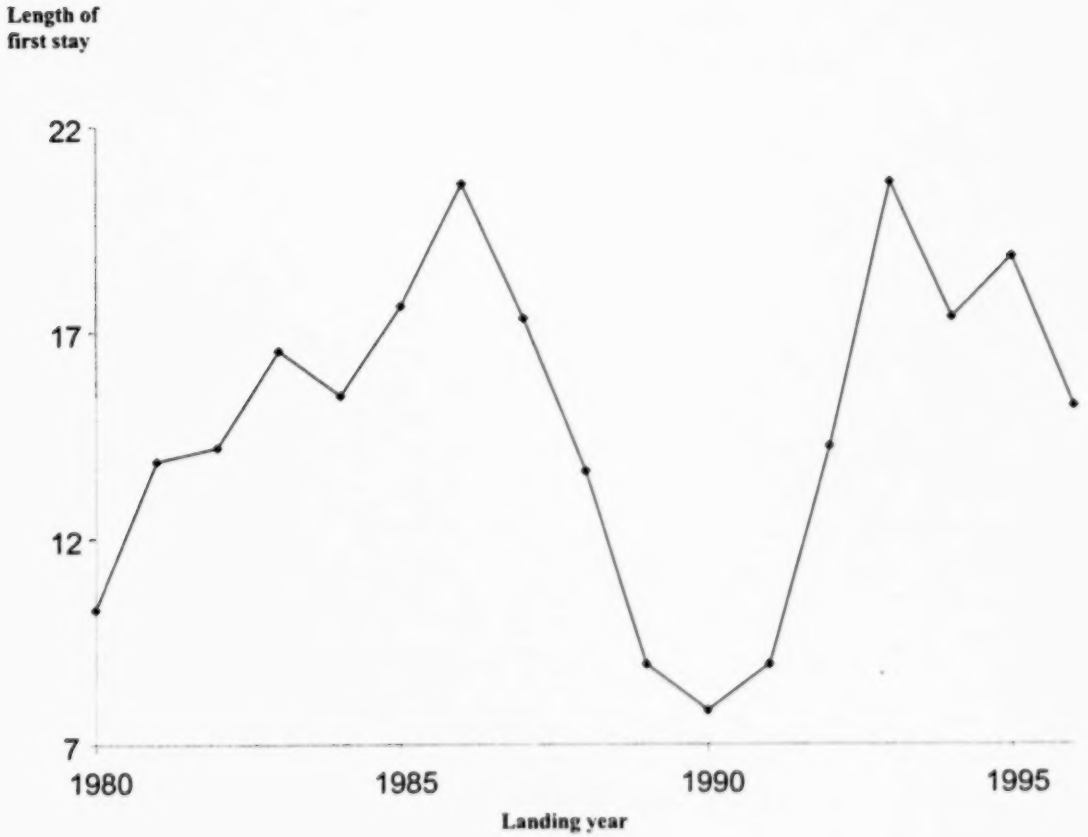
Source: Authors' calculation based on model estimates from Table 2.

Figure 3
The probability of a absence in the first 20 years by landing year



Source: Authors' calculation based on model estimates from Table 3.

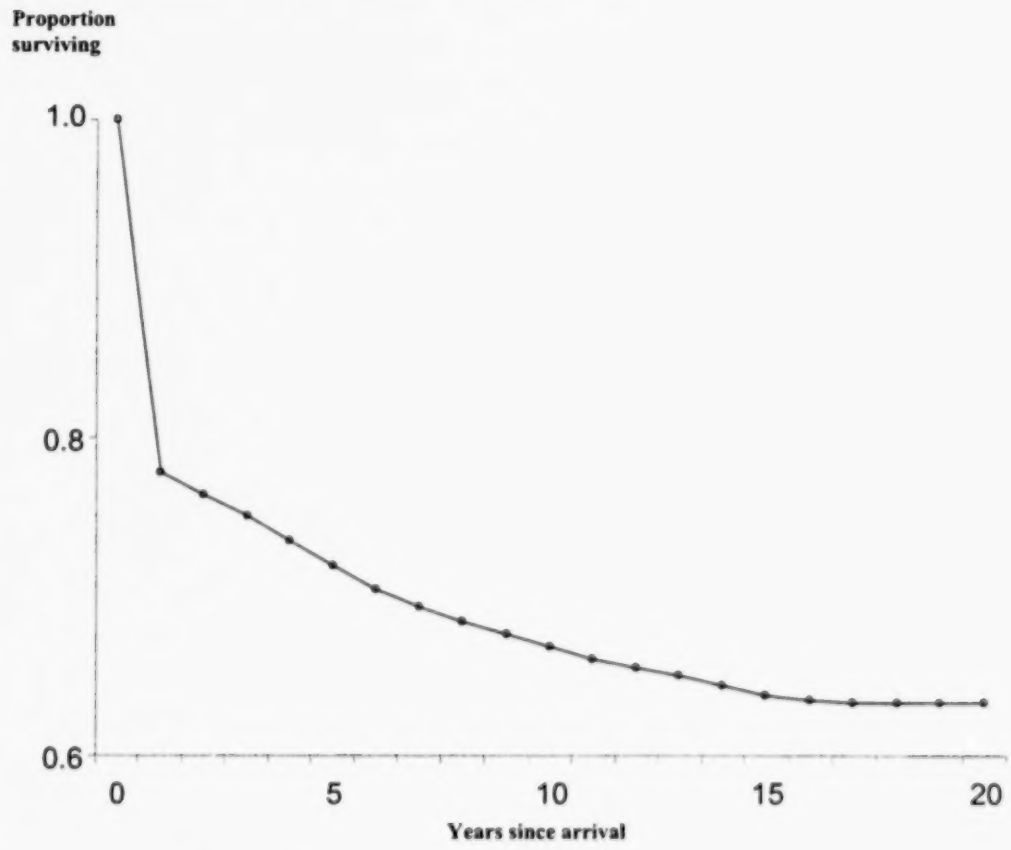
Figure 4
The expected length (in years) of the first spell by landing year



Note: The expected lengths of stay estimated in Table 3 are converted into years in this figure for expositional purposes.

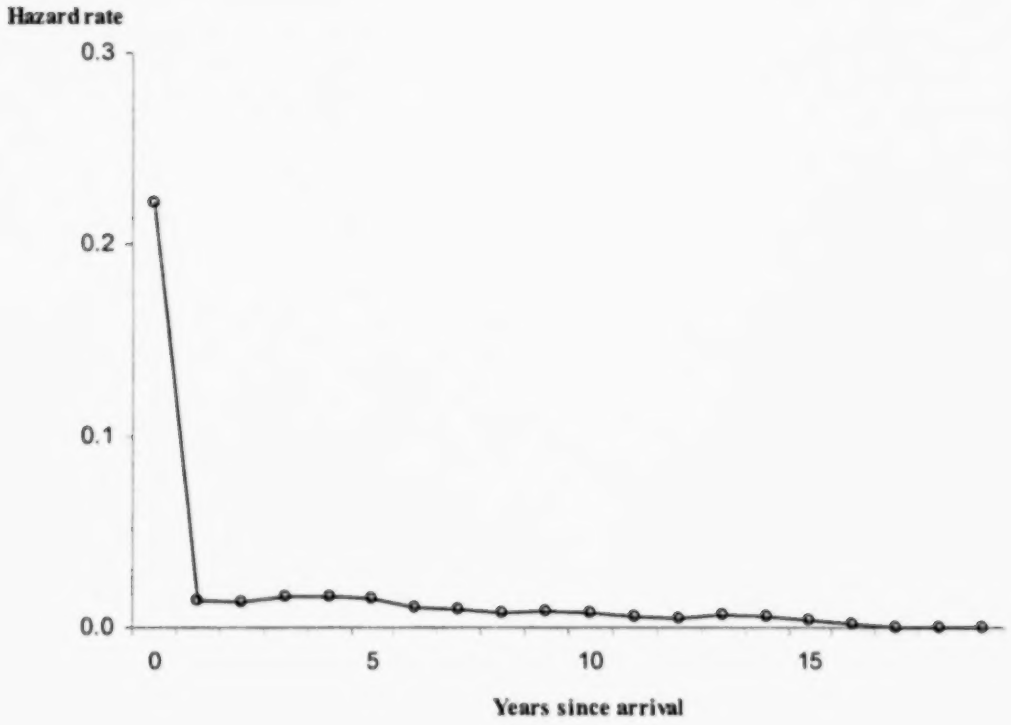
Source: Authors' calculation based on model estimates from Table 3.

Figure 5
Survival functions using life tables - All immigrants



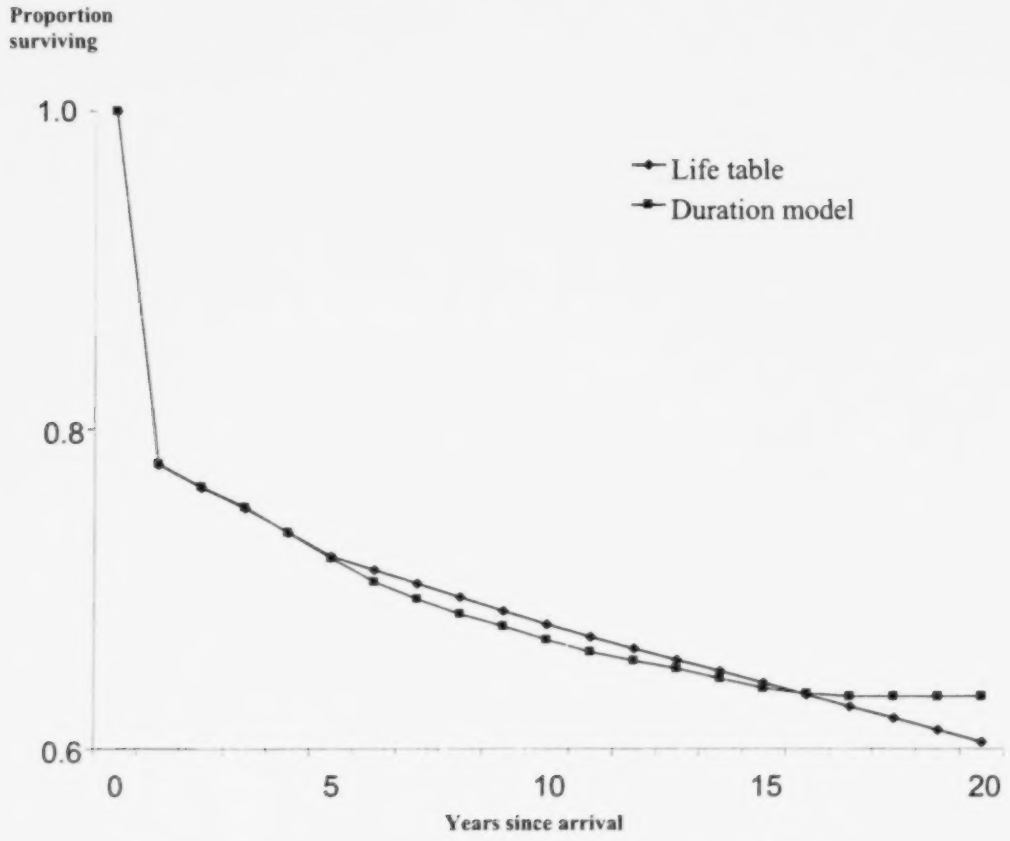
Source: Authors' calculation based on Life Table analysis.

Figure 6
Estimated hazard rate from life tables – All immigrants



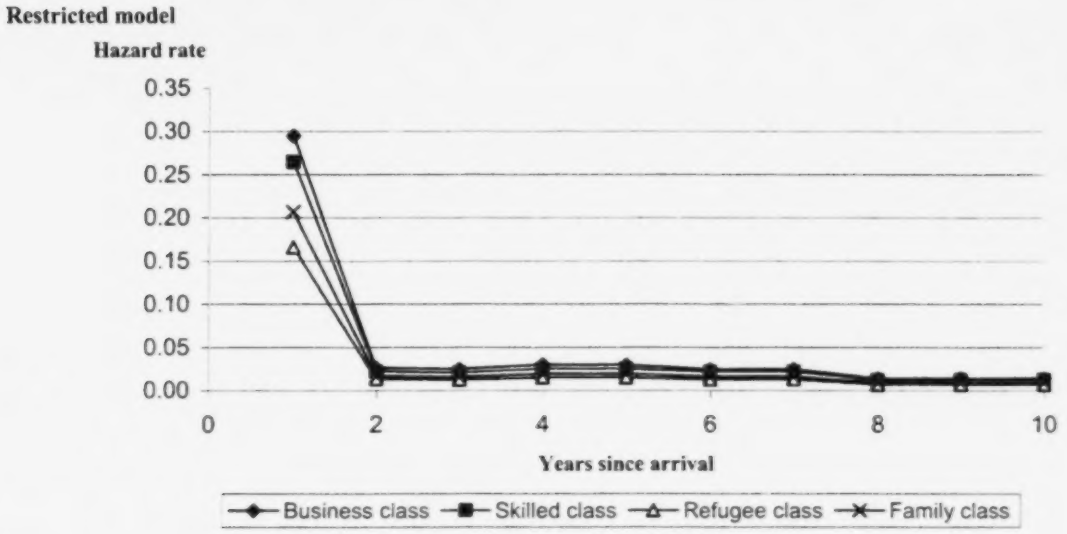
Source: Authors' calculation based on Life Table analysis.

Figure 7
Predicted and empirical survival functions



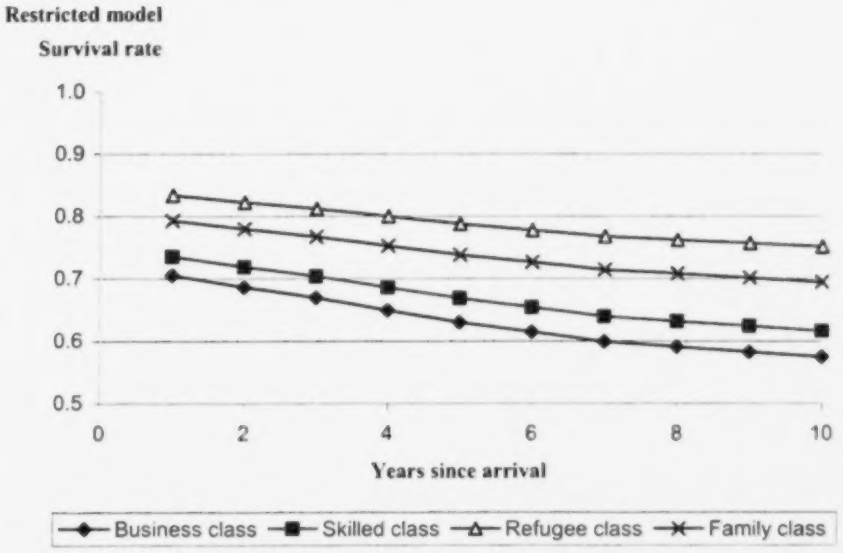
Source: Authors' calculation based on Life Table estimates and proportional hazard model.

Figure 8a
Discrete proportional hazard rates by visa class



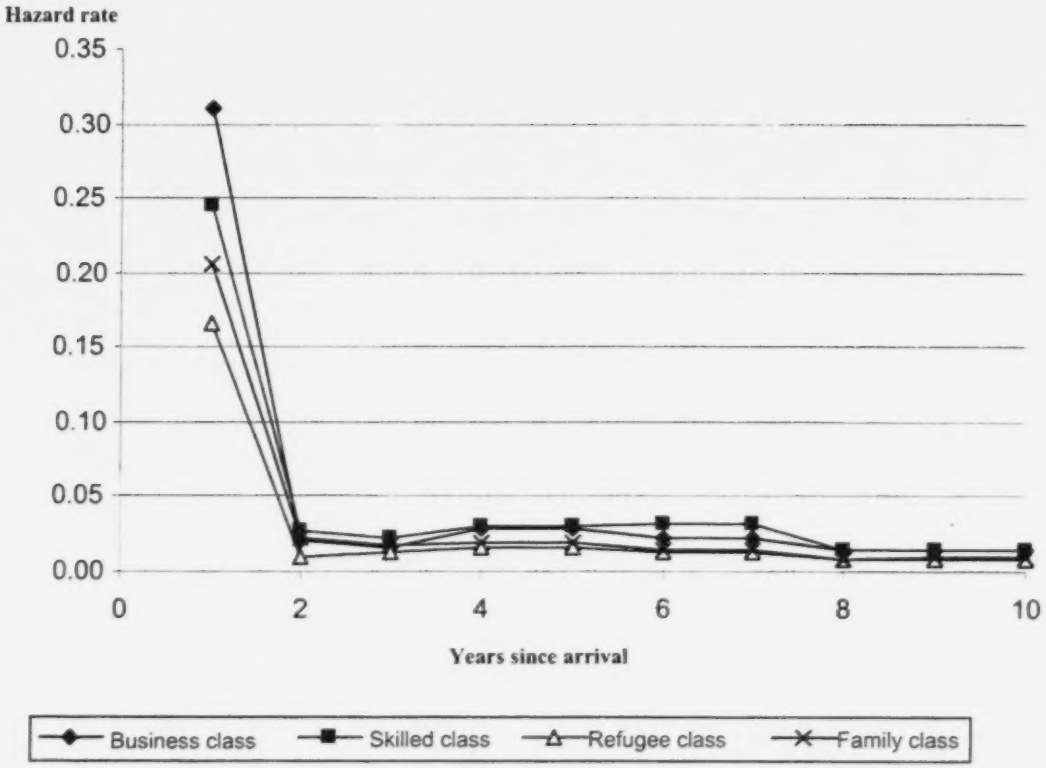
Source: Authors' calculation based on Duration model estimates in Table 4.

Figure 8b
Discrete proportional survival rates by visa class



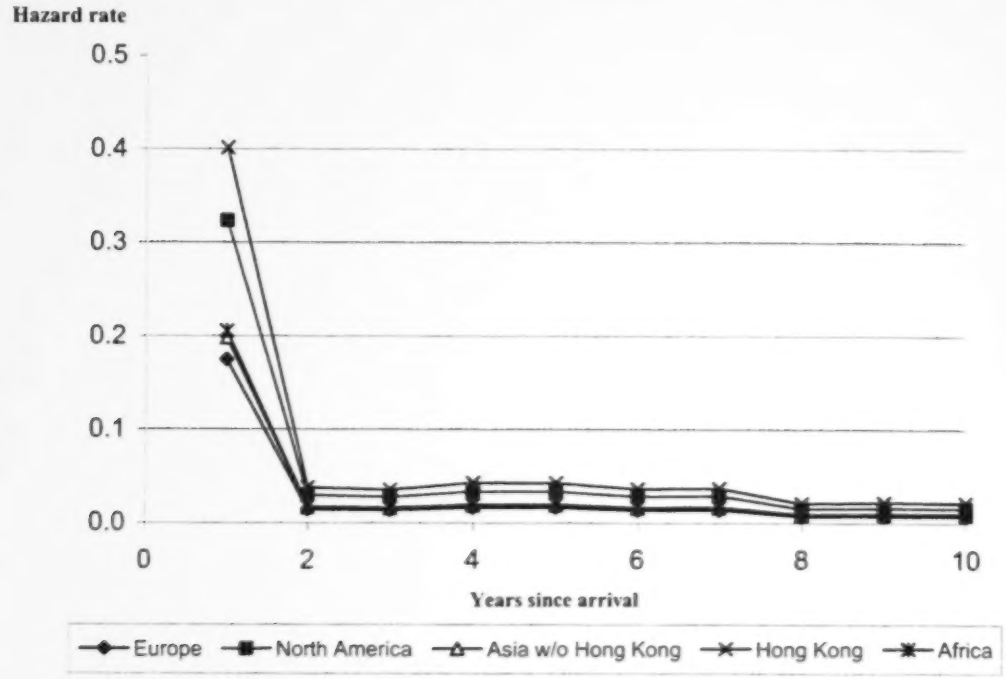
Source: Authors' calculation based on Duration model estimates in Table 4.

Figure 9
Discrete hazard rates by visa class



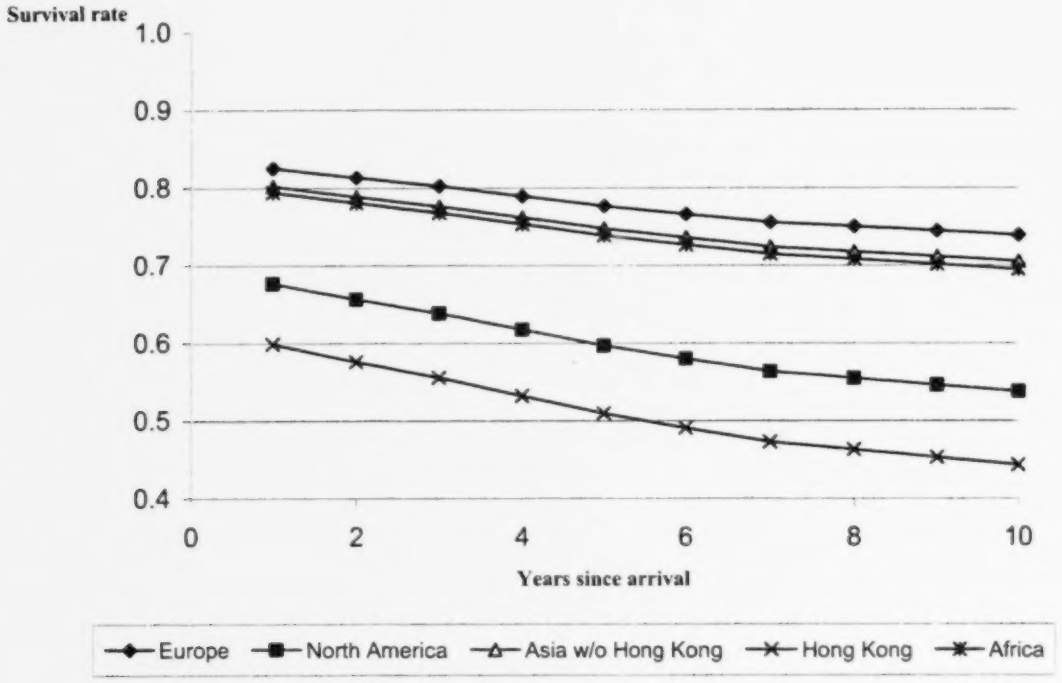
Source: Authors' calculation based on Duration model estimates by visa class.

Figure 10a
Discrete proportional hazard rates by source region



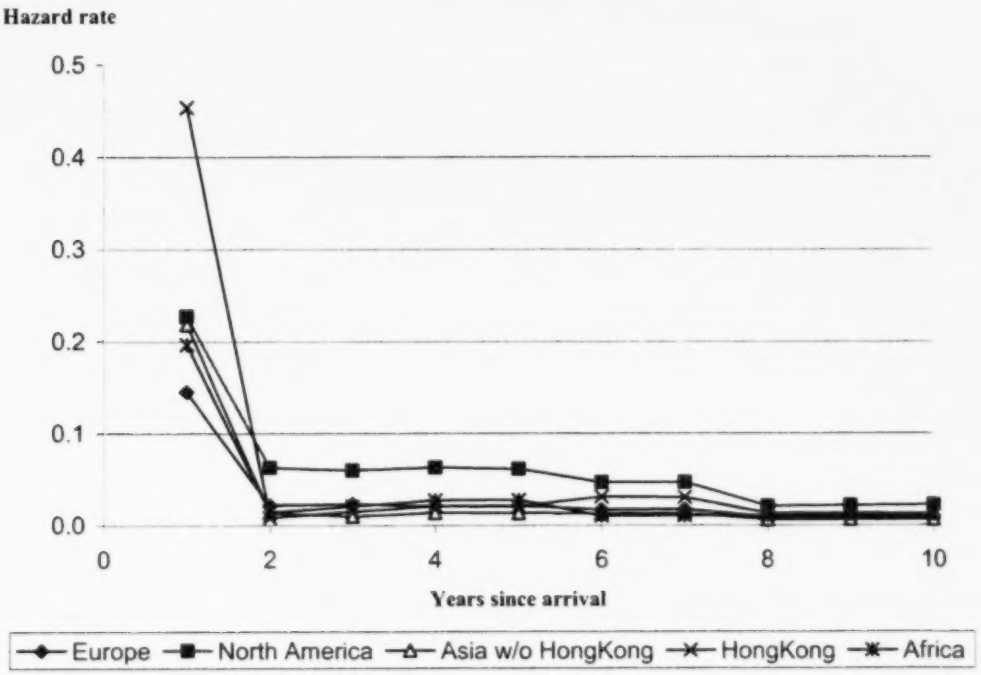
Source: Authors' calculation based on Duration model estimates in Table

Figure 10b
Discrete proportional survival rates by source region



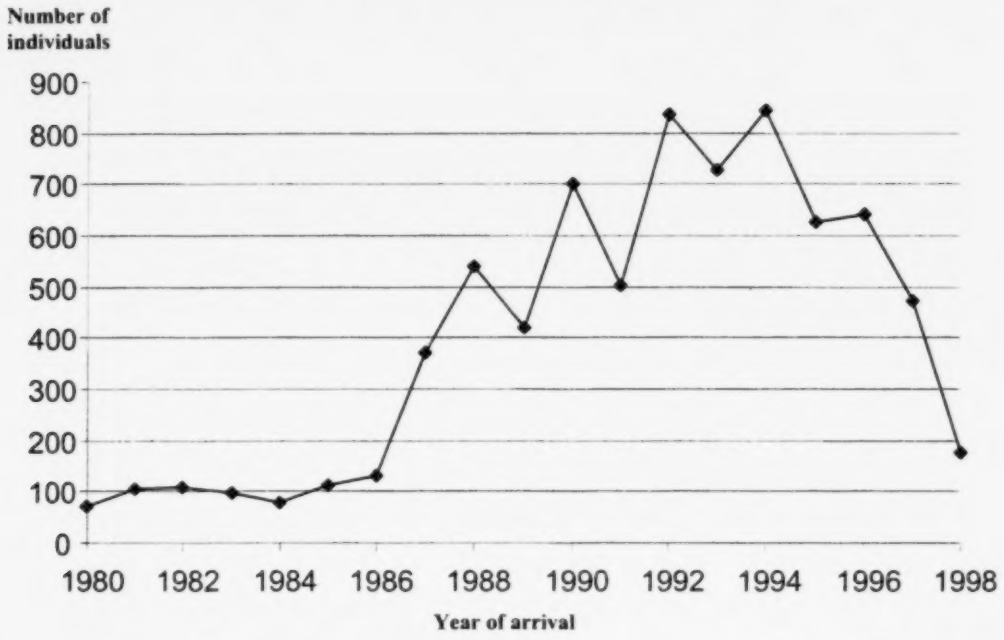
Source: Authors' calculation based on Duration model estimates in Table

Figure 11
Discrete hazard rates by source region



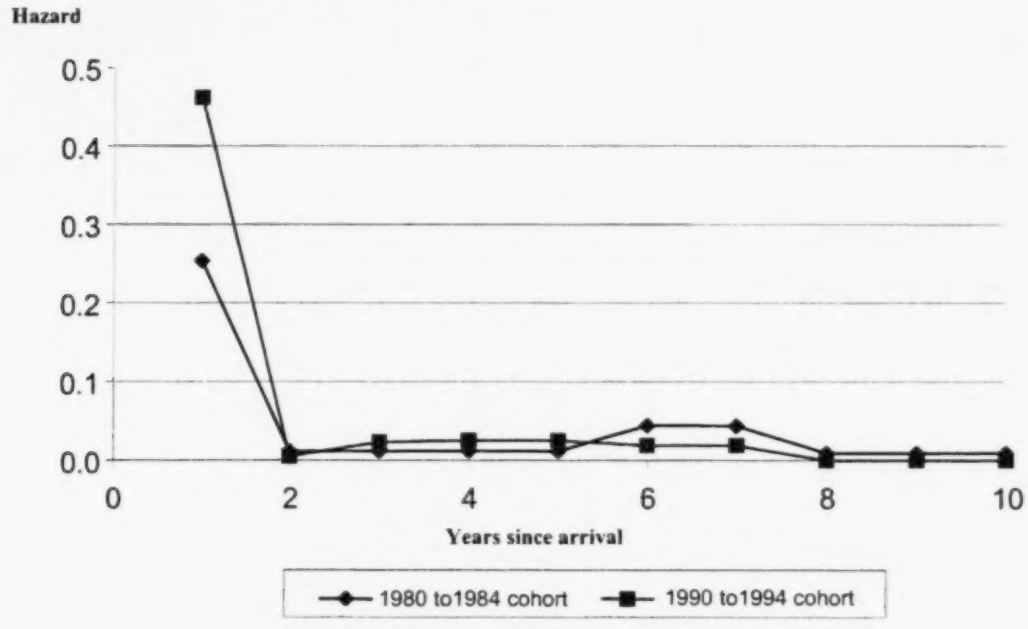
Source: Authors' calculation based on Duration model estimates by source

Figure 12
Landings of males aged 25 to 45 from Hong Kong, 1980 to 1996



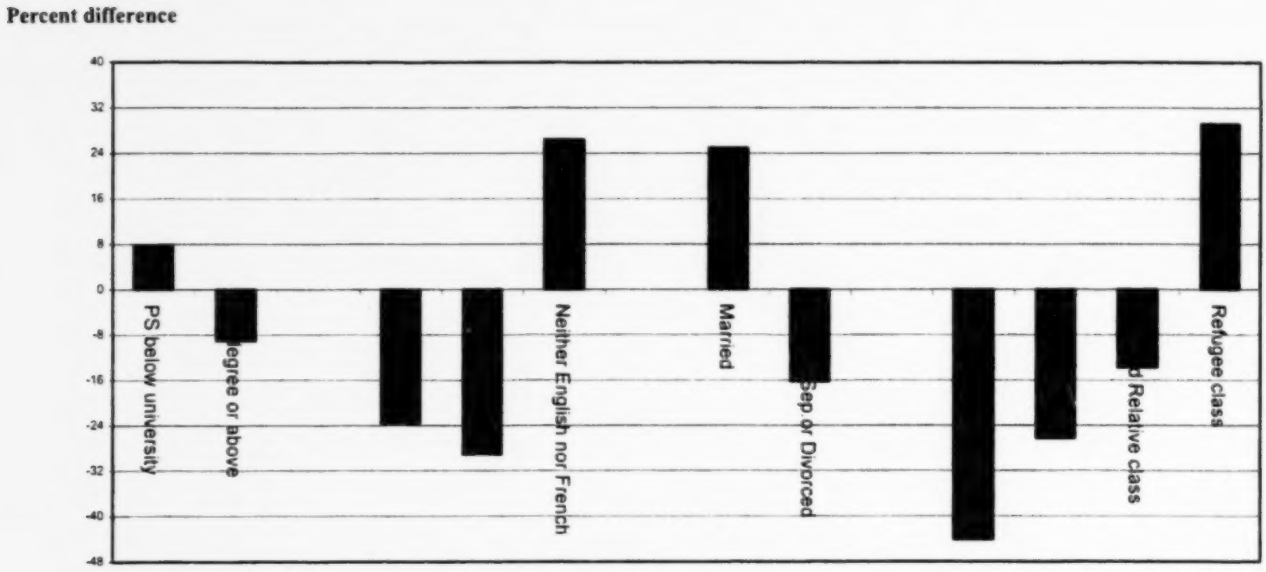
Source: Authors' calculation based on landings data.

Figure 13
Discrete hazard rates for Hong Kong by cohort



Source: Authors' calculation based on Duration model estimates for Hong Kong.

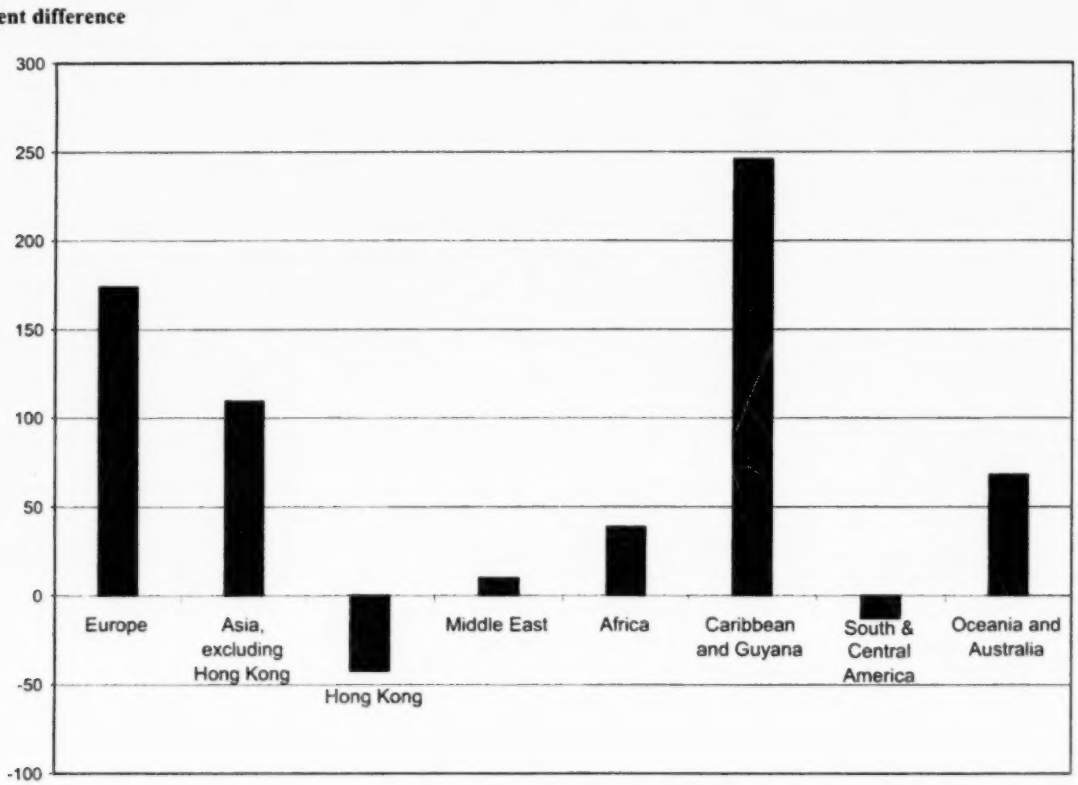
Figure 14
Percent differences in duration of stay by education, language ability and admission class



Note: The differences in this figure are with respect to a reference person who is a single individual, with education level below post-secondary level, fluent in English, admitted under family class. The bar chart for "Business class", for example, shows the difference in length of stay between a business class immigrant and the reference person who is a family class immigrant.

Source: Authors' calculations based on interval regression analysis results in Table 3.

Figure 15
Percent differences in duration of stay by region of origin



Note: Percent differences in the figures are with respect to an immigrant who arrived from North America.

Source: Authors' calculations based on interval regression analysis results in Table 3.

Appendix

Definition of spells

The spells are based on tax filing information. Individuals have an "opportunity" to file in any year that they are resident in the country. The indicator that a spell ended is four or more consecutive missed opportunities to file, ignoring the first opportunity on landing. Consider two individuals landing in January and December 1980, respectively. The first opportunity to file considered for the purposes of the spell definition is the tax year 1981.

If there is no filing in 1981, and no filing for the following 3 years, ending with the tax year 1984, then:

(i) the end date of the spell is an unknown date between January 1981 and December 1981 for the individual landing in December 1980, and between February 1980 and December 1981 for the individual landing in January 1980; and

(ii) the start date of the spell is the landing date, January or December 1980.

If there is filing in 1981, but no filing in the next four opportunities, ending with the tax year 1985 then:

(i) the end date of the spell is an unknown date between January 1982 and December 1982 for both individuals, and

(ii) the start date of the spell is the landing date January or December 1980.

Finally, if there is a re-appearance in the tax files, say in the 1987 and 1988 tax years, followed by 4 consecutive years of non-filing, ending with the tax year 1992, then:

(i) the end date of the second spell is an unknown date between January 1989 and December 1989 for both individuals, and

(ii) the start date of the spell is an unknown date between January 1987 and December 1987.

Figure A1
Tax filing rates by sex

Source: Authors' calculation based on Longitudinal Administrative Data File (LAD) coverage estimates.

References

Angelucci, M. 2003. "US Border Enforcement and the Inflow-Outflow of Mexican Illegal Migration." Mimeo.

Beach, C. M., A. G. Green and J. G. Reitz (eds.). 2003. *Canadian Immigration Policy for the 21st Century*. John Deutsch Institute for the Study of Economic Policy. McGill-Queen's University Press.

Borjas, G. J. and B. Bratsberg. 1996. "Who Leaves? The Outmigration of the Foreign-Born." *The Review of Economics and Statistics*. LXXVIII, 1: 165–176.

Colussi, A. 2004. "Migrants' Networks: An Estimable Model of Illegal Mexican Immigration." Mimeo.

Constant, A. and D. S. Massey. 2002. "Self-Selection, Earnings, and Out-Migration: A Longitudinal Study of Immigrants to Germany." Institute for the Study of Labour (IZA) Discussion Paper No. 672.

Jasso, G. and M. R. Rosenzweig. 1982. "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States." *Demography*. 19, 3: 279–290.

Jenkins, S. 2003. Survival Analysis with Stata Course Notes. Institute for Social and Economic Research (ISER), University of Essex.

Lam, K.-C. 1994. "Outmigration of Foreign-Born Members in Canada." *Canadian Journal of Economics*. XXVII, 2: 352–370.

Massey, D. S., J. Durand and N. J. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Free Trade*. New York: Russell Sage Foundation.

Meyer, B. D. 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica*. 58, 2: 757–782.

Michalowski, M. 1991. "Foreign-born Canadian Emigrants and Their Characteristics, (1981-1986)." *International Migration Review*. 25, 1: 28–59.

Munshi, K. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*. 118, 2: 549–599.

Piore, M. 1979. *Birds of Passage: Migrant Labor and Industrial Societies*. New York: Cambridge University Press.

Prentice, R. and L. Gloecker. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics*. 34, 1: 57–67.

Warren, R. and J. Marks Peck. 1980. "Foreign-Born Emigration From the United States: 1960 to 1970." *Demography*. 17,1: 71–84.